Metabolomic approaches for the identification of metabolic pathways in *Trypanosoma brucei*

Katharina Louise Johnston, Dipl.- Biol. Univ. MRes

Submitted in fulfilment of the requirements for the degree of a Doctor in Philosophy

School of Life Science
College of Medical, Veterinary and Life Science
University of Glasgow

March 2015

Abstract

Trypanosoma brucei is a parasitic protozoan that can cause human African trypanosomiasis (HAT) and Nagana in cattle. Human African trypanosomiasis is deadly when left untreated, and thus there is an urgent need to develop new drugs against this disease. As trypanosomes are early diverged eukaryotes, it is anticipated that studying their metabolism can identify novel drug targets. The main drug currently in use against the late encephalitic stage, Eflornithine, was shown to inhibit an essential pathway in trypanosomes (Yarlett and Bacchi, 1989).

In this Thesis three approaches were used to apply metabolomic and proteomic techniques for protein function identification and to investigate metabolic pathways. The genome of *T. brucei* has been published (Berriman et al., 2005) and data is available via databases, such as TriTrypDB, a database dedicated to the trypanosomatids (Aslett et al., 2009). An estimated 40% of the identified genes in this organism are annotated with an unknown or putative function. In 2006, Saito *et al.* developed a systematic method to ascertain enzyme function based on an *in vitro* assay, in combination with metabolite profiling. This approach was successfully applied in several other studies. Here, I investigate the use of this method for its application in a high throughput approach for unknown enzyme identification in trypanosomes. Seven putative identified enzymes were randomly selected from TriTrypDB, cloned and expressed in *E. coli* and a function could be attributed to at least one of the enzymes. Furthermore, the amino acid metabolism in trypanosomes was investigated; using stable isotope labelling combined with metabolomics. The flux of labelled compounds could be traced through the organism showing the active metabolic pathways of L-methionine, L-proline and L-arginine in *T. brucei*.

Two *T. b. brucei* strains used in this study, GVR35 and 427, cause different forms of infections in their mammalian host. GVR35 causes a chronic infection and invades the central nervous system (CNS) with varying parasitemia in mice, whereas infection with strain 427 presents an acute form with high parasitaemia, causing high mortality, without invading the CNS. What causes this difference in the progression of infection? Secreted or excreted proteins from the parasites, referred to as the secretome, have been described as being important factor for virulence and avoiding the host immune response (Geiger *et al*., 2010) and Garzon *et al*. (2006) showed that excreted/secreted proteins can inhibit the maturation of dentritic cells and stop them from inducing a lymphocytic allogenic

response. Significant differences in proteins secreted from these two strains are discussed; although the results are preliminary.

Table of Contents

List of Tables

List of Figures

Acknowledgement

First and foremost, I would like to thank my supervisors, Prof. Mike Barrett, Dr.Richard Burchmore and Dr. Karl Burgess for their help and support throughout my PhD. You introduced me to the field of metabolomics and proteomics and I am very grateful for the opportunity to experience this. Especially Mike, who showed me, that even as an expert you can still get excited about the endless possibilities this research holds.

Fiona Achcar has been a great teacher since moving into the GBRC and without her patience and helpfulness, the last two years would have been a lot less enjoyable. A special thanks to Fiona also for our discussions and her superb efforts proofreading my thesis.

I would also like to acknowledge the Doctoral Training Centre and the funding bodies BBSRC and EPSRC for the opportunity to do this PhD.

Hyun Kim, you have been a great friend and were always reliable running my samples on the mass spec – through good columns and through bad columns.
Alan Scott for all your help with my protein work, you went out of your way to help me and proteomics and protein purification would have not been possible without you!
Snezhana, you also have been a great support during my PhD, helping me with my proteomic experiments. But more importantly you have become a wonderful friend.
Dave Wildridge and Lesley Morrison, thank you for your help and support with the molecular side of my project. You have been a great support over the last three and a bit years, not only when advice was needed, but you also kept me calm when things were not going the way I wanted them to.
To all the members of the Barrett group, past and present. Alan, Fede, Gordon, Isabel, Jane, Julie, Kevin, Pieter and Roy. Thank you for all your help and advice over the last years.
The Level 5 students who I shared an office with: Alan, Suleman, Shafquat, Becca, Taylor, Daniel, Lola, Joanne and Harshal, you will be greatly missed.
My parents, my sister and her husband Peter, thank you for your encouragements and support, even from a distance! Not to forget the daily, early phone calls to get me out of bed during the write up period, just a reminder, Germany is one hour ahead of Scotland!
Last, but not least, a big thank you to my husband Dave for sticking with me during these challenging times.

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, this dissertation is the result of my own work and has not been submitted for any other degree at The University of Glasgow or any other institution.

Katharina Johnston

Abbreviations

ABC	ammonium bicarbonate
Acetyl-CoA	acetyl - coenzyme A
ADP	adenosine diphosphate
AMP	adenosine monophosphate
ATP	adenosine triphosphate
bp	base pair
BLAST	Basic Local Alignment Search Tool
BBB	blood brain barrier
BSA	bovine serum albumin
BSF	bloodstream form
cAMP	cyclic adenosine monophosphate
CE	capillary electrophoresis
CHAPS	(3-[(3-Cholamidopropyl)-dimethylammonio]-propane- sulfonate)
CMM	Creek's minimal medium
CMW	Chloroform/Methanol/Water
CNS	central nervous system
°C	degrees Celsius
CoA	coenzyme A
2DE	two dimensional gel electrophoresis
DFMO	difluoromethylornithine
dH_2O	distilled water
DiGE	Differential Gel Electrophoresis

DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dSAM	decarboxylated S-adenosylmethionine
DTT	dithiothreitol
EC	enzyme commission
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetra acetic acid
EtOH	ethanol
ESI	electrospray ionisation
FA	trifluoroacetic acid
FAD	flavin adenine dinucleotide
FMN	Flavin mononucleotide
FBS	fetal bovine serum
FT-MS	Fourier Transform Mass Spectrometry
g	gram
GC	gas chromatography
GC-MS	gas chromatography mass spectrometry
gDNA	genomic DNA
GDP	guanosine diphosphate
GTP	guanosine triphosphate
h	hour
HAT	human African trypanosomiasis
HPLC	high pressure liquid chromatography
HILIC	hydrophilic interaction chromatography

HMI-9	Hirumi medium 9
IDEOM	identification of metabolites
IEF	isoelectric focussing
IMAC	immobilised metal ion affinity chromatrography
IPG	Immobilized pH gradient
IPTG	isopropyl-β-D-thiogalactoside
kb	kilobase
kDa	kilodalton
kDNA	kinetoplast DNA
KEGG	Kyoto Encyclopedia of Genes and Genomes
L	litre
LB	Luria-Bertani medium
LC	Liquid Chromatography
LC-MS	liquid chromatography mass spectrometry
M	molar
MeOH	methanol
MOPS	3-(N-Morpholino)propanesulfonic acid
MS	mass spectrometry
MTA	Methylthioadenosine
mRNA	messenger RNA
µg	microgram
µl	microlitre
µM	micromolar
mg	milligram

ml	millilitre
mM	millimolar
min	minutes
MTA	methylthioadenosine
nm	nanometre
nM	nanomolar
NAD^+	nicotinamide adenine dinucleotide
NADH	nicotinamide adenine dinucleotide (reduced form)
$NADP^+$	nicotinamide adenine dinucleotide phosphate
NADPH	nicotinamide adenine dinucleotide phosphate (reduced form)
NMR	nuclear magnetic resonance
NTDs	neglected tropical diseases
OD	optical density
ODC	ornithine decarboxylase
ppm	parts per million
PAGE	poly-acrylamide gel electrophoresis
PBS	phosphate buffered saline
PCA	principle components analysis
PCR	polymerase chain reaction
PP	pyridoxal phosphate
PPP	pentose phosphate pathway
rpm	revolutions per minute
RNA	ribonucleic acid
RNAi	RNA interference

s	second
SAM	S-adenosylmethionine
SDS-PAGE	SDS polyacrylamide gel electrophoresis
SDS	sodium dodecylsulphate
TbAK	*Trypanosoma brucei* arginine kinase
TCA	tricarboxylic acid
Tet	tetracycline
TFA	trifluoroacetic acid
TIC	total ion current
TPP	thiamine pyrophosphate
UDP	uracil diphosphate
UMP	uracil monophosphate
UTR	Untranslated Region
VSG	variable surface glycoprotein
V	volt
WHO	World Health Organisation
wt	wild type
ZIC-HILIC	zwitterionic hydrophilic interaction chromatography

Chapter 1

1. Introduction

1.1. Human African trypanosomiasis

Human African trypanosomiasis (HAT), also known as sleeping sickness, is a parasitic disease endemic to sub-Saharan Africa. The disease is fatal when untreated and it affects mainly rural populations.

HAT belongs to the neglected tropical diseases (NTDs), a group of 17 diseases affecting mainly the world's poorest population in 149 countries. Included in those are infections caused by protozoan parasites, leishmania, Chagas disease and HAT, as well as others caused by bacteria (like leprosy), viruses (including dengue and rabies) and helminths (e.g. onchoceriasis and schistosomiasis). The World Health Organisation (WHO) estimates that about 1.4 billion people are affected by at least one of those diseases (http://www.who.int/neglected_diseases/diseases/en/). As those NTDs only affect populations in underdeveloped countries, little resources are spent to develop drugs or vaccines.

In the case of HAT, the WHO estimated that about 70 million people are at risk of acquiring this disease in sub-Saharan Africa. While the risk for the population varies depending on the location, five million people live in high risk areas for HAT (Figure 1.1) (Franco et al., 2014). Several major HAT outbreaks have been reported since the late 19th century and, although HAT was thought to be under control in the mid-20th century, HAT re-emerged in the 1990s due to political instabilities in the affected regions (Barrett, 2006; Brun et al., 2010). Recent figures show the numbers of reported cases of HAT dropping below 10,000 since 2009 after being as high as over 30,000 in the late 1990s (Franco et al., 2014).

HAT is caused by an infection with subspecies of the protozoan parasite *Trypanosoma brucei*. The parasite is transmitted to the human host by the bite of the tsetse fly (Genus *Glossina*). As it is a vector borne disease, the occurrence of HAT is limited to the distribution area of the tsetse fly.

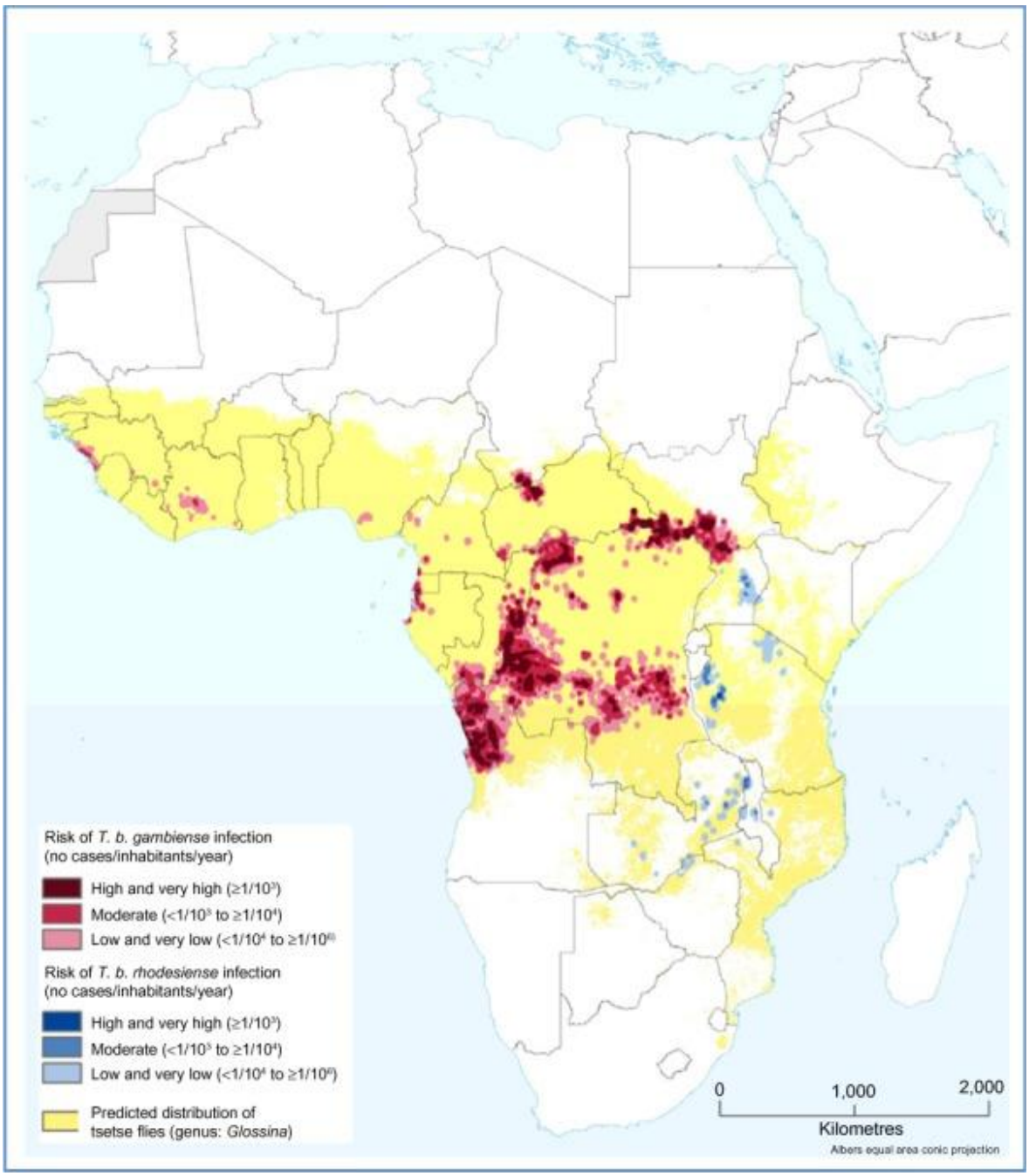


Figure 1.1: Population at risk from infection of HAT. HAT caused by *T.b.gambiense* is marked in red and infections caused by *T.b.rhodesiense* in blue. Distribution area of the tsetse fly is highlighted in yellow **(Simarro et al., 2012)**.

The occurrence of HAT in Africa can be divided in two areas: Central-west Africa, where *Trypanosoma b. gambiense* causes a chronic form of HAT and east Africa for *Trypanosoma b. rhodesiense* causes an acute form of HAT. Only in Uganda are both subspecies present. After the parasites get transmitted by the tsetse fly they proliferate at the site of infection, causing an inflammatory nodule. This trypanosomal chancre rarely occurs in infections caused by *T. b .gambiense*, but in 50 % of infections with *T.b. rhodesiense* (Barrett et al., 2003).

The disease manifests itself in two stages:
Stage 1, also early haemolymphatic stage, starts when the parasites spread from the trypanosomal chancre to the lymph nodes and enter the bloodstream. Multiple organs might also be infected (Barrett et al., 2003; Kennedy, 2004). Early symptoms of HAT are non-specific, including episodes of fever, malaise and headaches. A typical symptom for HAT is lymphadenopathy, which develops in *T. b. gambiense* infections after several weeks. Slave traders in the 18th century used neck swellings as signs for this infection, as described by Thomas Masterman Winterbottom (Barrett et al., 2003). Stage 2, also called late encephalitic stage, starts when the parasites invade the central nervous system (CNS). In infections caused by *T. b. rhodesiense* this happens within a few weeks, while it could occur between several months and years when infected with *T. b. gambiense*. Symptoms during second stage of HAT include disturbances to sleeping patterns, which gives the disease its name of 'sleeping sickness'. Other symptoms can include mental and psychiatric disturbances (Rodgers, 2009). Untreated HAT will lead to coma and eventually death. Although treatment of early stage HAT is relatively effective (Rodgers, 2009), HAT becomes more symptomatic during the second stage.

There are currently four licensed drugs for the treatment of HAT, which can be administered depending on sub-species and stage of the disease. The drugs for the treatment of stage 1 HAT are suramin and pentamidine. Suramin was first used against HAT in 1922 (Miézan et al., 1994) and can be used against both sub-species. Pentamidine, a drug from the early 1940s, is only effective against *T. b. gambiense*. Both drugs are easy to administer and are also relatively safe compared to their counterparts for stage 2 treatment (Barrett et al., 2007). However, due to HAT becoming more symptomatic during the second stage of the disease safer drugs are urgently needed.

Once the parasites invade the CNS the drugs available are melarsoprol and eflornithine.
Melarsoprol was first synthesised in 1949 and acts against both *T. b. gambiense* and *T. b. rhodesiense*. However, it has severe side effects that can result in the death of about 5% of the patients (Blum and Burri, 2002). Eflornithine, a drug developed as a potential antineoplastic agent in the 1970s, is recommended by the WHO for the treatment of HAT caused by *T. b. gambiense*. Recently a combination therapy of eflornithine and nifurtimox (a drug used for the treatment of Chagas disease) has been developed and has been shown to cure 98% during clinical trials (Priotto et al., 2007).

Of all the drugs available for HAT treatment, only eflornithine has a known mode of action (Poulin et al., 1992; Vincent et al., 2012)

1.2. Trypanosomes

Trypanosomatida is a group of kinetoplastid protozoa that are exclusively parasitic and are mainly found in insects. However, a few do have a two host life cycle and can infect a wide range of hosts, from vertebrates to invertebrates and plants. Species of this order known to infect humans are members of the genus *Leishmania*, with approximately 12 million people in South America, Middle East and India infected (Singh et al., 2012). *Trypanosoma cruzi*, causing Chagas disease in South America, affecting an estimated 10 million people (Barfield et al., 2011) and *Trypanosoma brucei*, causative agent of sleeping sickness in Africa (Barrett et al., 2003). Trypanosomes are ubiquitous parasites that can cause disease not only in humans but also in animals. They occur mainly in Africa; however, species are also known to cause human disease in South America and other animal disease outside Africa (Gibson, 2007).

The species *Trypanosoma equiperdum*, which causes dourine in equines, is thought to be closely related to *T. evansi*, (causative agent of surra in equines). Both *T. equiperdum* and *T. evansi* have a wider distribution than other trypanosome species, being reported in Africa, most of Asia, Russia, parts of the Middle East, South America and southeastern Europe ((Gibson, 2007); Animal Health Information Database). *T. equiperdum* is thought to be the only member of the trypanosome family to be exclusively sexually transmitted.
Trypanosoma cruzi, as mentioned above, is only found in South America and causes Chagas disease in humans.

1.2.1. *Trypanosoma brucei*

There are three subspecies of *T. brucei*, two of which are infectious to humans and are associated with the disease HAT.
T. b. gambiense, responsible for 98% of all cases, causes a chronic form of HAT in Central and West sub-Saharan Africa.
T. b. rhodesiense, accounts for 2% of HAT, causing an acute infection in East sub-Saharan Africa.

T. b. brucei causes nagana in cattle and other wild and domestic animals; however it fails to infect humans due to its sensitivity to an innate immune complex found in human serum. There are two trypanolytic factors (TLF) found in human serum, TLF1 and TLF2, causing lysis of trypanosome parasites. TLF1 is a component of high density lipoprotein (Hajduk et al., 1994) and TFL2 is an apolipoprotein-A1/IgM complex (Tomlinson et al., 1995). The lytic component in both is Apolipoprotein L-1 (APOL1)(Vanhamme et al., 2003). Lysis of the parasites occurs after uptake of the TLFs into the parasite when APOL1 is taken up into endosomal and lysosomal membranes, causing osmotic swelling and lysis by forming cation selective pores in membranes (TLF1) (Molina-Portela et al., 2005). The human infective subspecies *T. b. gambiense* and *T. b. rhodesiense* show resistance against lysis by human TLFs. In *T. b. gambiense* reduced TLF binding or uptake was described due to reduced expression of HpHbR (Kieft et al., 2010). In *T. b. rhodesiense* resistance of APOL1 is achieved by expression of a serum resistance-associated protein (SRA).

1.2.2. Life cycle

Trypanosoma brucei has a two host life cycle; an insect host (tsetse fly) and a mammalian host (see Figure 1.2). The two subspecies infectious to humans, *T. b. gambiense* and *T. b. rhodesiense*, can also infect wild animals. Although they do not fall ill, animals can serve as a reservoir for these parasites. This is particularly true for *T. b. rhodesiense* (Enyaru et al., 2006); while *T. b. gambiense* is more reliant on human to human transmission via the tsetse fly (Brun et al., 2010).

Trypanosomes get transmitted during the blood meal of the tsetse fly. Transmition to the mammalian host is in form of metacyclic trymastigotes present in the saliva of the tsetse fly. In the mammalian (or human) host, metacyclic trypomastigotes differentiate into bloodstream trymastigotes (long, slender form) and proliferate at the site of infection for a few days, forming a trypanosomal chancre. From there, the fast dividing bloodstream form trypomastigotes spread to the lymph nodes and the bloodstream where they can also infect multiple organs. The long slender bloodstream form rapidly multiplies and the parasitemia in the blood increases. The parasites then differentiate into non-dividing short stumpy trypomastigotes which can be taken up by the tsetse fly during a blood meal. In the midgut of the tsetse fly the parasites differentiate into procyclic trypomastigotes and proliferate. Epimastigotes leave the midgut and differentiate into metacyclic trypomastigotes,

migrating to the salivary gland where they can be transferred to the mammalian host when the tsetse fly takes another blood meal.

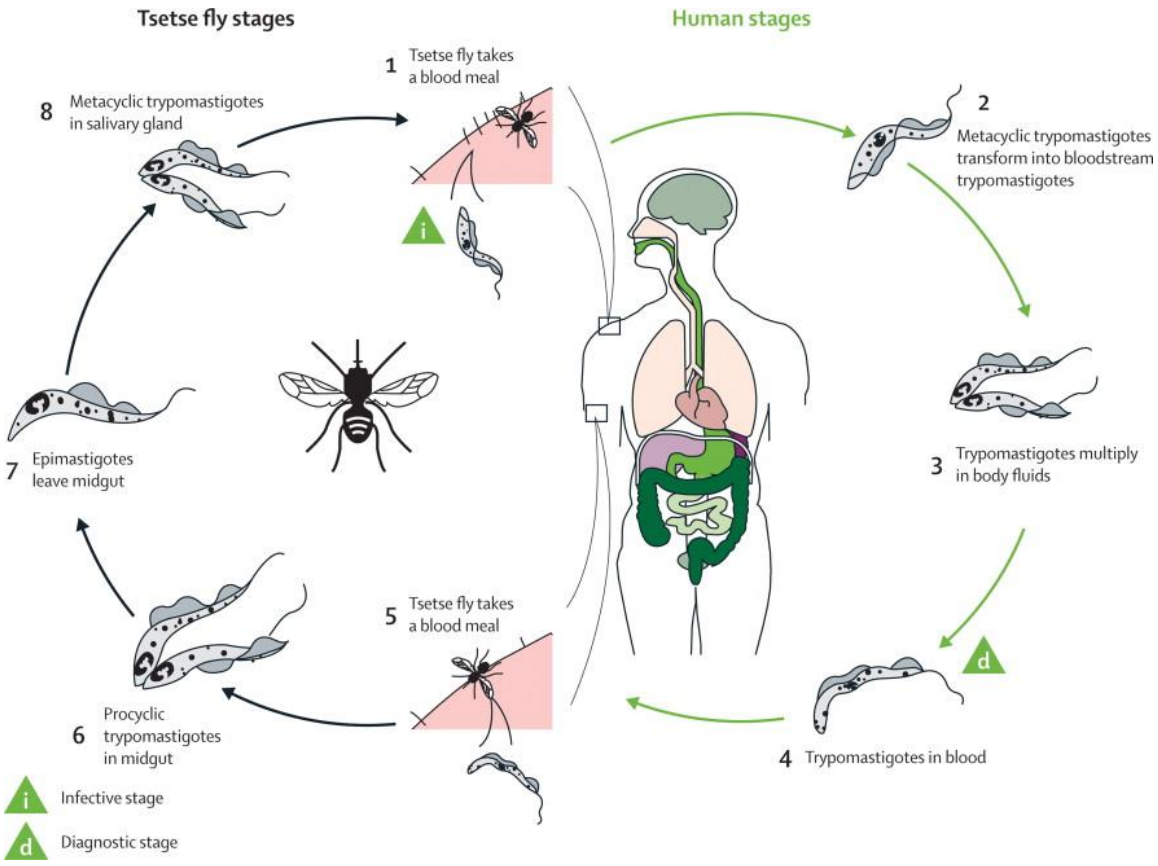


Figure 1.2: Representation of the two host life cycle of *Trypanosoma brucei* in the tsetse fly and human host (Blum et al., 2008). Permission to reproduce this image has been granted by Elsevier.

1.2.3. *Trypanosoma brucei* strains (used in research)

From the three subspecies of *Trypanosoma brucei* only two cause the human disease sleeping sickness. However, the human infectious strains *T. b. gambiense* and *T. b. rhodesiense* share >99% genome sequence identity with *T. b. brucei* (Jackson et al., 2010) making the non-infectious form a safe model organisms to use in research.

Trypanosoma brucei brucei strains commonly used in the laboratory environment:

- 'TREU 927' originated in Kiboko, Kenya, is a pleomorphic strain that was used for the whole genome sequencing program (Berriman et al., 2005) Although the strain 427 is more commonly used in laboratories the ability of TREU 927 to complete the whole two host life cycle made it a better candidate for the sequencing project (Peacock et al., 2008).

- ‘427’ or ‘Lister 427’ is a monomorphic strain, which lost its ability to differentiate into short stumpy forms reportedly due to frequent passage in the lab (Peacock et al., 2008). There are mixed information in the literature regarding the origin of the ‘commonly used 427’ strain. ‘Lister 427’, a *T.b.brucei* strain, was isolated from cattle in former Tanganyika (now Tanzania) in 1956. ‘s427’, is thought to be *T.b.rhodesiense*, isolated from sheep in 1960 (South East Uganda). (http://tryps.rockefeller.edu/trypsru2_pedigrees.html)
- GVR35, originally isolated from wildebeest in the Serengeti (1966), causes a chronical infection in mice and when infected mice are left untreated the parasite manifest itself in the CNS within 21 days. This makes it a good model organism for studying trypanocidal drugs for the late encephalitic stage of this disease (Frevert et al., 2012; Jennings et al., 2002)

1.3. Biology of trypanosomes

Trypanosomes are considered to be very early diverged eukaryotes and make very good model organisms for studying biological processes. But what makes trypanosomes good model organisms? First, reliable culture methods exist for both life cycle stages. Host-pathogen interaction can also be studied in suitable animal models - tsetse fly for the insect stage and rodents, mice or rat, for the bloodstream form. There is a wide range of molecular techniques available for trypanosomes; gene knockout by homologous recombination can be performed, as well as RNA interference for gene knockdown (Alsford and Horn, 2008). Trypanosomes also have some very distinctive features and organelles, making them an interesting model to study and compare biological processes in eukaryotes (Smith and Bütikofer, 2010). A few biological processes initially described in trypanosomes are antigenetic variation for evading the immune system (Cross, 1975), glycosylphosphatidylinositol (GPI) anchoring for VSGs (Ferguson and Williams, 1988), trans-splicing (Sutton and Boothroyd, 1986) and RNA editing (Benne et al., 1986). Their unique method of evading the human immune system through antigenetic variation has been extensively studied and uncovered several other aspects about this parasite, such as mechanisms of exocytosis. The genome, sequenced in 2005, is organised differently from other eukaryotes and they have a streamlined metabolism, consistent to their parasitic lifestyle.

1.3.1. Genome of trypanosomes

The nuclear genome of *T. brucei* is composed of three classes of chromosomes, which are distinguished by size. It contains 11 pairs of diploid megabase chromsosomes, which have been sequenced and published (Berriman et al., 2005), a variable number of intermediate chromosomes (IC) and between 50-100 mini-chromosomes (MC) (Melville et al., 1998).
The megabase chromosomes contain housekeeping genes as well as telomeric VSG expression sites (Melville et al., 2000). Approximately 9,000 genes were discovered, with around 900 pseudogenes. Around 1,700 genes are thought to be *T. brucei* specific genes (Berriman et al., 2005). In most *T. brucei* strains the copy number of ICs are 1-7 and the size varies between 200-700 kb, while the MCs are more numerous and between 30-150 kb (Wickstead et al., 2004). Both ICs and MCs play a major role in the recombination of bloodstream form VSGs and therefore the parasites ability to avoid the human immune system. While the ICs contain VSG expression sites, MCs contain non-transcribed basic copies of VSG genes (Melville et al., 1998).

The results of whole genome sequencing project has been made available from several sources, the most commonly used is the trypanosomatidae specific database TriTrypDB, which grants easy access to genome annotations (Aslett et al., 2010). Recently '-omics' datasets have been added to this database.

1.3.2. Evading the immune system

The bloodstream form *T.brucei* is a solely extracellular parasite, exposing them to the mammalian immune system. Therefore trypanosomes need to have a mechanism in place to avoid destruction. Variant Surface Glycoproteins (VSG) are expressed by the parasites, during their life cycle in the mammalian host, which form a dense surface coat covering the cell (Englund et al., 1982). Each trypanosome expresses only one VSG gene at any given time from a telomeric expression site (Horn and McCulloch, 2010). Those VSGs can be detected by the host's immune system and antibodies are produced against those. However, trypanosomes use antigenic variation of their surface coat to change the composition of those VSGs, which allows some parasites to evade the immune system (Horn, 2004).

1.3.4. Organelles

Trypanosomes are single celled flagellates, with an elongated cell shape which is defined by their microtubule cytoskeleton (Sherwin and Gull, 1989). Their organelles differ in some aspects to most eukaryotic cells (Figure 1.3). Morphological features of this parasite include two areas of DNA, one the nucleus and the other the kinetoplast, giving the group kinetoplastidae (which trypanosomes belong to) their name. The kinetoplast is a DNA containing region at the posterior end of a single elongated mitochondrium which is another striking morphological feature of trypanosomes (Kilgour, 1980).

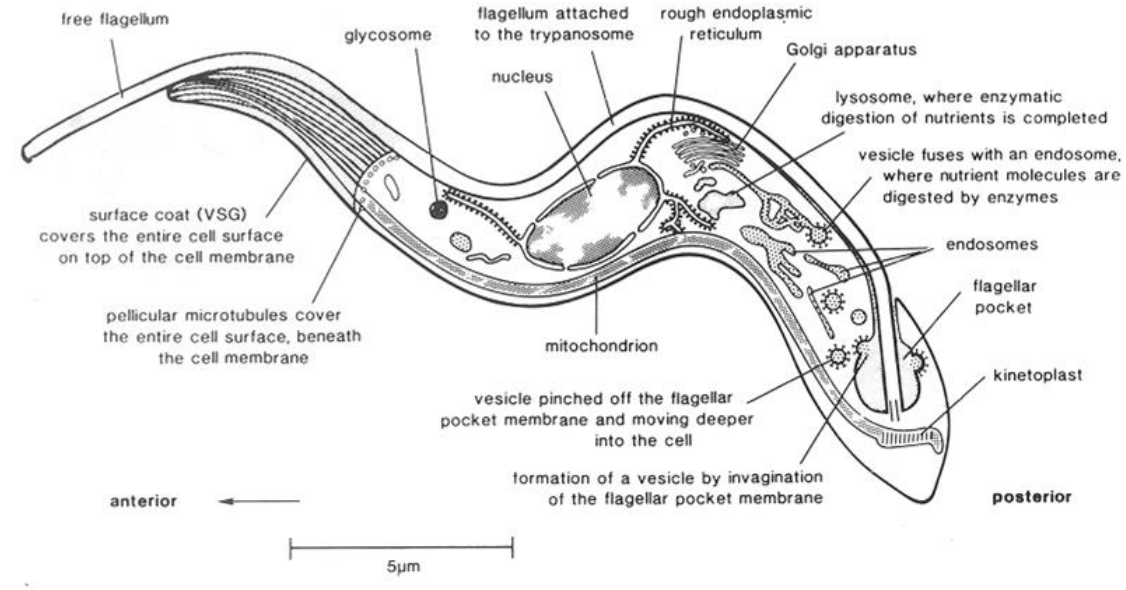


Figure 1.3: Schematic diagram of bloodstream form *T. brucei* illustrating major organelles. This picture originally appeared in ILRAD Reports. Vol7 (1), January 1989 and is available online http://www.ilri.org/InfoServ/Webpub/fulldocs/Ilrad90/Trypano.htm

The kinetoplast accounts for 10-20% of the cell's DNA and is arranged in mini and maxi circles (Stuart and Gelvin, 1982). The maxi circles (~30-50 copies/cell) contain genes encoding for some mitochondrial proteins, whereas the mini circles (~10,000 – 50,000 copies/cell) encode for short guide RNAs which are used as templates for the post-transcriptional editing of maxi circle transcripts (Simpson and Shaw, 1989). The single mitochondrion stretches from the posterior to the anterior end of the cell. The form and function of the mitochondrion differs between the different life stages of trypanosomes. In bloodstream form the mitochondrion lacks cristae, folds of the inner membrane, reflecting the absence of mitochondrial respiration (Matthews, 2005).

At the posterior end of the trypanosome lies the flagellar pocket, the only site for endo – and exocytosis in trypanosomes (Overath and Engstler, 2004). Also, the flagellar pocket is the exit point for the flagellum, which is an essential organelle for the viability of trypanosomes (Kohl et al., 2003). The flagellum is not only the sole means of motility, but also plays an important factor in development, transmission and pathogenesis (Langousis and Hill, 2014).
A unique feature to trypanosomatidae is the glycosome, an organelle where part of the glycolysis takes place.

1.3.5. Metabolism

Due to their complex life cycle (switch between mammalian and insect host), trypanosomes must have the ability to adapt quickly to their changing environment. This is not only important for the developmental cell biology (Matthews, 2005) but also applies to the changes in metabolism between procyclic and bloodstream form cells. Also, trypanosomes show a reduced metabolism coinciding with their parasitic life style (Nerima et al., 2010). They lack pathways for purine biosynthesis and until recently it was thought that they rely solely on the host for components such as fatty acids and sterols as well (Fairlamb, 1989). However, it was shown that trypanosomes are capable of de novo synthesis of fatty acids (Smith and Bütikofer, 2010).

Lipids make between 11-18% of the dry weight of *T. brucei*, with a distribution similar to the range of lipids found in other eukaryotes, namely phospholipids, neutral lipids, fatty acids, isoprenoids and sterols (Smith and Bütikofer, 2010). After discovering that the fatty acid molecules of the glycosylphosphatidylinositol (GPI) anchors of the VSGs are exclusively myristate (Ferguson and Cross, 1984), it was speculated that bloodstream form trypanosomes could indeed synthesize myristate as there would be a higher demand on myristate from the parasite than provided in the blood (Lee et al., 2007). Trypanosomes mainly synthesise fatty acids through an endoplasmic reticulum (ER) based elongase (ELOs) pathway instead of type I or type II fatty acid synthesis (Stephens et al., 2007). Type II fatty acid synthesis pathway does exist in the mitochondrion, but is not the main source of myristate (Lee et al., 2006). In bloodstream form trypanosomes myristate is the endpoint of fatty acid synthesis (Smith and Bütikofer, 2010), while procyclics produce stearate (Stephens et al., 2007).

Energy metabolism in trypanosomes has been well studied in both life cycle stages. It differs from its host's, due to localisation of glycolytic enzymes in organelles called glycosomes (Opperdoes and Borst, 1977), reduced metabolic pathways and the regulation of glycolytic enzymes (Nwagwu and Opperdoes, 1982). Energy metabolism also differs between the two life stages of the parasite. Some of the features are described below.

The most unusual aspect of the trypanosome energy metabolism is the compartmentalisation of the first nine glycolytic enzymes in the glycosomes. In bloodstream form 90% of the proteins found in the glycosomes are glycolytic enzymes (Aman et al., 1985). Although the glycosomes were named after presence of glycolytic enzymes, other metabolic pathways are also localised within those organelles. Pathways (other than glycolysis) described to be localised within the glycosomes include the pentose phosphate pathway (PPP), β-oxidation of fatty acids, purine salvage and biosynthetic pathways for pyrimidines, ether-lipids and squalenes, although activity varies between trypanosomatids (Michels et al., 2006).

In bloodstream form trypanosomes, glycolysis is the only source of ATP production and pyruvate the endproduct of glycolysis (van Hellemond et al., 2005). In contrast, procyclic trypanosomes can use L-proline as energy source, which is more readily available in the midgut of the tsetse fly than glucose (Bursell, 1963) and, to a lesser extent, L-threonine (Bringaud et al., 2006); however, they do prefer D-glucose when available (Lamour et al., 2005). But even when glucose is used in procyclics, the energy metabolism differs as is seen by the end products, which are succinate, acetate and alanine (Cazzulo, 1992). It has been shown in cultured trypanosomes that they are able to adapt very quickly to changes in nutrition offered and they can switch between L-proline metabolism and D-glucose metabolism within one hour (Coustou et al., 2008).

The redox metabolism also differs from their mammalian host as trypanosomes rely on the thiol trypanothione, a metabolite unique to trypanosomes. Trypanothione is produced via the conjugation of two glutathione molecules with one spermidine (Fairlamb et al., 1985). The metabolites playing a role in the synthesis of trypanothione are putrescine, the precursor to spermidine, from L-ornithine and the formation of spermidine from decarboxylated S-adenosylmethionine. Enzymes involved in polyamine and trypanothione biosynthesis are essential for parasite growth in *T. brucei* (Willert and Phillips, 2012) and the identification of pathways involved have gained a lot of research interest due to their potential in drug discovery. The drug eflornithine for example, which is currently in use

against stage 2 HAT, is a suicide inhibitor of the enzyme ornithine decarboxylase, blocking the formation of putrescine from L-ornithine leading to the death of the parasites. Formation of decarboxylated S-adenosylmethionine from S-adenosylmethionine is regulated by the enzyme S-adenosylmethionine decarboxylase. The human S-adenosylmethionine decarboxylase is stimulated by putrescine. However the trypanosome enzyme is allosterically activated by the formation with a catalytical dead paralogue, termed prozyme, and is not affected by putrecine (Willert et al., 2007). The appearance of inactive enzyme homologues (prozymes) has been observed in a variety of enzyme families during a genome search of metazoan species (Pils and Schultz, 2004). In trypanosomes so far three enzymes have been found where prozymes increase the function of enzymes, namely hexokinase (Chambers et al., 2008; Morris et al., 2006), deoxyhypusine syntethase (Nguyen et al., 2013) and the above mentioned S-adenosylmethionine decarboxylase.

1.4. Metabolomics

Metabolomics is an emerging field of postgenomic biology that aims to identify and quantify small cellular metabolites (Mr < 1200) within a given biological sample.

The ‘-omics’ technologies comprises genomics for identification of the genome, transcriptomics for studying gene expression, proteomics for the analysis of all proteins present in the cell and metabolomics. Transcriptomics, proteomics and metabolomics are considered more dynamic fields than genomics as they can vary in response to environmental conditions (van der Werf et al., 2005), with metabolomics being described as the closest representation to the phenotype. Originated as recently as the early 1990s, Metabolomics is considered the latest contribution to the ‘-omics’ technologies. The rising interest in metabolomics in recent years can be seen in the rise in the number of publications in the area. In 1999 three publications with the keyword metabonomic/metabolomic could be found, but this number reached 203 in 2004 (Dettmer et al., 2007). In a similar search using the medline trend website (Web resource at URL:http://dan.corlan.net/medline-trend.html) with the terms “Metabolomics” or “Metabonomics” a similar rise in publications could be found from the years 2000 to 2013 (see Fig 1.4).

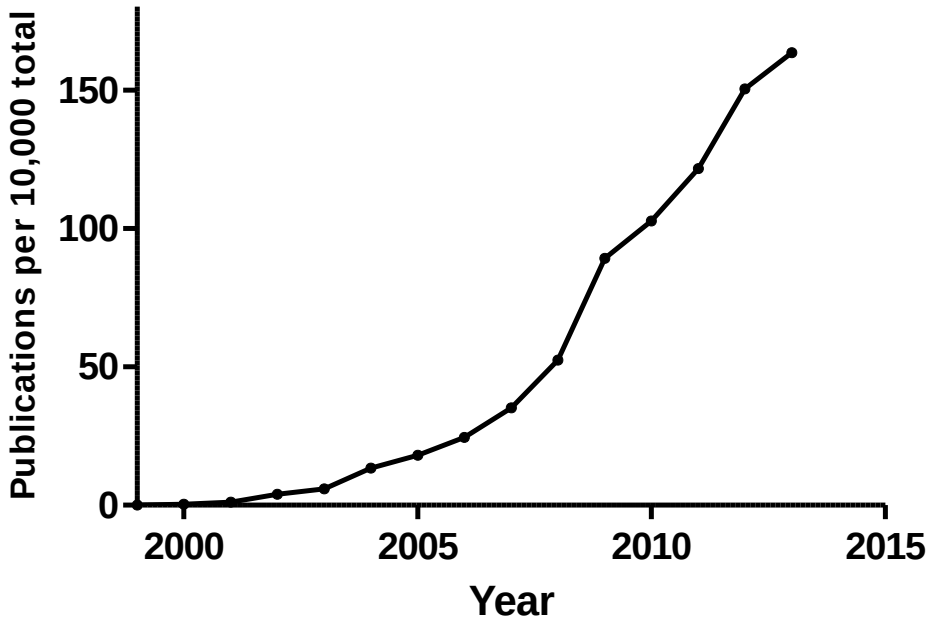


Figure 1.4: Medline trend created for publications using the term "Metabolomics" or "Metabonomis". Publication number was normalised against the total number of publications in that year. Trend was obtained from: Alexandru Dan Corlan. Medline trend: automated yearly statistics of PubMed results for any query, 2004.

However, the rising number of publications can also be explained as a 'knock on effect', as more technology and method development in the field of metabolomics are being made available, the more it becomes applicable to a wider range of research areas. Metabolomics is increasingly applied to investigate microbial, plant, environmental and mammalian systems (Brown et al., 2011). However, the wide range of chemical diversity and range of concentrations (dynamic range) of different metabolite classes makes it difficult to produce a single analytical platform. Metabolites represent a more diverse set of chemical diversity when compared to proteins (constituted by the 20 amino acids) and the nucleic acids and therefore provides wide variations in chemical and physical properties (Dunn and Ellis, 2005). Consequently, it is necessary to use a variety of techniques for complete coverage of all cellular metabolites (Castle et al., 2006).

1.4.1. Nuclear magnetic resonance spectroscopy (NMR spectroscopy)

NMR was one of the first techniques to be used for a broad untargeted metabolite profiling since the 1970s (Beckonert et al., 2007). Many NMR-based applications for metabolomics have been published but NMR has been commonly utilized for biomarker discovery by investigating numerous diseases and toxic processes studying body fluids (Nicholson et al.,

2002). The main benefits for using NMR include highly reproducible quantitation even across different instruments and simple sample preparation without the need for derivatization (Creek et al., 2012a). Limitation of NMR includes the relatively poor sensitivity compared to mass spectrometry (Creek et al., 2012a).

1.4.2. Mass spectrometry (MS)

MS is a well established analytical technique that measures the mass to charge ratio of charged particles and has been established as one of the most essential tools for metabolomic research (Baran et al., 2009). The principle of mass spectrometry consists of ionizing chemical compounds to create charged molecules (or fragments). These charged molecules can then be detected by analysing the mass to charge ratio. A MS instrument can be divided into three different parts; an ion source, a mass analyser, and a detector. As the "-omics" approaches require the analysis of complex mixtures, novel analytical methods / techniques had to be introduced. A variety of mass spectrometry based approaches are commonly used, such as mass spectrometry coupled to liquid chromatography (LC-MS), gas chromatography (GC-MS) and capillary electrophoresis (CE-MS) with different types of mass spectrometers (Barrett et al., 2010).

1.4.2.1. Chromatography

Direct injection of samples into the mass spectrometer (MS) has the advantage that no metabolites are lost during sample preparation. However, in complex samples and samples with high salt content ion suppression can occur, as well as adduct formation during the electrospray process (Dettmer et al., 2007). Combining chromatography with mass spectrometry can reduce ion suppression as complex samples are separated before entering the MS. Which chromatography system to use depends on the kind of molecules studied, the two most popular systems are Gas chromatography-MS (GC-MS) and Liquid chromatography-MS (LC-MS).

Gas chromatography-MS is often described as the gold standard for metabolomics (Garcia and Barbas, 2011). Ion suppression of co-eluting compounds rarely occurs, compared to LC-MS (Koek et al., 2011). Other benefits of using GC-MS are the capacity for high quantitative accuracy and reproducibility of samples, as well as the use of 'libraries' to identify compounds by their mass spectra (Schauer et al., 2005). As samples have to travel through the column in gas form with temperatures around 350°C, compounds suitable for

the GC-MS platform are low molecular weight (<350Da), volatile and thermally stable metabolites (Dunn, 2008). However, compounds can be made volatile and thermally stable by chemical derivatisation, and here lies the drawback of GC-MS. Sample preparation can be complex and lead to multiple derivatisation products. Furthermore, sample stability is not guaranteed (Dunn and Ellis, 2005).

Liquid chromatography-MS is the more prominent technology due to ability for greater coverage of the metabolome (Garcia et al., 2008) and is becoming increasingly popular for the use in metabolomics. Sample preparation is less complex than GC-MS as samples do not need to be derivatised. Samples go through the column in liquid phase and are separated according to their polarity. Reversed-phase LC-MS is well established in metabolomics. However, many polar or charged metabolites are not retained. Therefore, another form of LC, hydrophilic interaction liquid chromatography (HILIC) with the introduction of robust and reproducible stationary phases has recently gained popularity in metabolomics studies (Cubbon et al., 2010). More recently approaches to combine reversed phase colums with HILIC in a single run has been successfully performed and it was shown, with the example of beer samples, that a greater coverage of metabolites was achieved (Haggarty et al., 2015).

1.4.2.2. Ionisation

Soft ionisation techniques have been the milestone in applying mass spectrometry to biological studies, such as metabolomics and proteomics. Electron impact (EI) ionisation is mainly applied to GC-MS studies. The spectra obtained contain many fragments which can be used for compound identification as fragmentation patterns are very reproducible (Watson, 2010). Electrospray ionisation (ESI) is commonly coupled to LC-MS. Ions are formed by transferring the solvents through a capillary, which is held at high voltage. Ions enter the mass analyser after exiting the capillary and transferred from liquid into gas state. ESI can be performed in both positive and negative mode, which gives a wider coverage of the metabolome (Dettmer et al., 2007).

1.4.2.3. Mass analysers

The mass analyzer is used to sort the ions by their masses by applying electromagnetic fields. A detector measures the mass to charge ratio (m/z) of the ions and provides data for calculating the abundances of each ion present.

For the analysis of biomolecular samples, the Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) and the Orbitrap are currently able to achieve high resolution and mass accuracy in the 1 ppm range (Breitling et al., 2006a). In FT-ICR-MS, trapping of ions is achieved by a strong magnetic field which confines the charged particles to a circular path. The Orbitrap, which was exclusively used in this study, is a powerful mass spectrometer that can examine a variety of types of chemical systems which makes it an appropriate analysing platform for studies such as proteomics and metabolomics. The Orbitrap was invented by Alexander Makarov as a new type of mass spectrometer and patented in 1999. It uses dynamic trapping of ions in an electrostatic field (Makarov, 2000), more specific orbital trapping based on the Kingdon trap (Kingdon, 1923). The trap consists of an outer barrel like electrode and a central spindle like electrode along the axis. Ions are injected tangentially into the electric field between the electrodes and trapped. The electrostatic attraction to the inner electrode is balanced by centrifugal forces which makes the ions cycle around the central electrode in rings. Additionally, the ions also move back and forth along the z-axis of the central electrode. The main distinction of the Orbitrap is that the mass to charge ratio is derived from the frequency of the ion oscillations along the axis of the field which can then be determined using image current detection and fast FT algorithms (Makarov, 2000). The image current is detected on split outer electrodes and then amplified by a differential amplifier. Orbitraps have a high mass accuracy (1–2 ppm), a high resolving power (up to 200,000) and a high dynamic range (around 5000) (Hu et al., 2005; Makarov et al., 2006).

1.5. Applications for metabolomics

The term metabolomics can be identified as the identification and quantification of all small molecules within a biological system (Dettmer et al., 2007). However, to date there is not a single analytical platform that can achieve this. The metabolome (which is identified as the complete set of all metabolites within an organism) represents a broad variety of compound classes and analytical platforms can be biased against one or more of them. Therefore, different approaches are in use to make the most of metabolomics technologies and depending on the research question asked. Below are the most common strategies used (Dettmer et al., 2007; Dunn, 2008; Dunn and Ellis, 2005).

Metabolite footprinting, also described as the exometabolome, is the analysis of the uptake of extracellular metabolites and the secretion of intracellular metabolites. The samples can be taken non-invasively.

Metabolite fingerprinting, also defined as intracellular metabolome, provides a snapshot of the global metabolism. Aims to sort samples from different biological origin or status intro groups rather than get a good identification and quantification of metabolites.

Metabolite profiling, aims for an untargeted analysis of the metabolome within a biological sample. To obtain a wide coverage of the metabolome multiple analytical platforms (or extraction methods) can be used.

Metabonomics, aims to show quantitative changes to the metabolome in response to pathophysiological stimuli or genetic perturbations.

Metabolome is the definition of the complete set of all metabolites present in an organism.

Stable isotope labelling, combined with metabolomics has gained an enormous research interest over the last few years. With the use of ^{13}C labelled compounds in metabolomics it is possible to track the carbon flux through a cell and get a much better inside of the metabolic pathways than with metabolomics alone.

The distribution of labelled carbons gives an indication of what metabolic pathways are active, new metabolites (or pathways) can be discovered.

The simplest way is to divide studies into two different approaches: targeted or untargeted. The targeted approach already requires knowledge about the metabolite(s) of interest and this approach is useful to gain knowledge of metabolite quantification or information about specific metabolic pathways. The untargeted approach is described as hypothesis generating (Creek et al., 2011).

Although terminology of metabolomics techniques used varies between different research groups, metabolomics is increasingly applied to the study of microbial, plant, environmental and mammalian systems (Brown et al., 2011). Applications include biomarker discovery, studies of gene function and the effects of drug therapy and toxicity (Nicholson et al., 2002).

1.6. Metabolomics and enzyme function identification

The full functional annotation of all genes (or gene products) within an organism displays a great challenge in the post-genomic era. Although the number of whole genome sequences available is rising, even in a well described organism like *E. coli* 40% of the genes have no confirmed function (Keseler et al., 2009). In other organisms the number is even higher. In the tritryps (*T. brucei*, *T. cruzi* and *Leishmania major*) this number reaches up to 70% combining the unknowns in all three species, although their genome sequences were published nearly 10 years ago (Choi and El-Sayed, 2012). With the data available from the whole genome sequencing projects, the potential function of a hypothetical protein can be determined by bioinformatics approaches, such as BLAST or Pfam domain searches. An alternative approach is the prediction of a protein's function based on the structure. This method is well described by (Whisstock and Lesk, 2003)which gives an overview of automated structure-based function prediction. However, it has been shown in the past that these annotations can be incorrect. In a study, Barrett *et al.* found that the trypanosome enzyme Tb927.8.2020, which had been putatively annotated as an arginase, did not show classic arginase activity (Hai et al., 2015; Vincent et al., 2012).

One of the tasks in functional genomics is to search for novel enzymes but also to provide functions for genes where bioinformatics can only give putative function. Metabolites can be used to assign or validate the function of enzyme activities as these are often the final downstream products or substrates of gene expression (Baran et al., 2009). Enzymatic reactions play an important role, not only in the regulation of all processes of life but also as biocatalysts in industrial processes or as targets in the discovery of new drugs (Liesener and Karst, 2005).

Traditional approaches to establish enzyme activity include enzyme assays, but the main drawback to this technique is that the enzyme function has to be known (Liesener and Karst, 2005). Integrated 'omics' approaches can be used to facilitate the identification of unknown enzymes (Fridman and Pichersky, 2005), however, prior biochemical knowledge was still essential. Mass spectrometry based analysis of small molecular metabolites has become a valuable method for the discovery and validation of functional assignments for enzymes (Baran et al., 2009).

1.6.1. Enzyme function identification by *in vitro* assay with recombinant protein

A systematic method to discover novel enzymes based on an *in vitro* assay in combination with metabolite profiling was introduced by Saito *et al.* (2006). Proteins of interest were purified and used for an *in vitro* reaction with a complex metabolite mixture. Bacto yeast extract, Bacto malt extract and Nutrition Broth were tested as a suitable metabolite source for their ease of preparation and low cost while still containing a large variety of compounds. The advantages of using natural and complex metabolite mixtures also include the possibility for screening of unknown type of reactions and an environment closer to the physiological conditions than standard *in vitro* essays (Saito et al., 2006). After the reaction the mixture was purified and analysed by CE-MS. The metabolite profiles of the reaction mixture incubated in the presence and absence of the protein of interest were compared. The idea is that the compound whose level decreases following incubation with the protein can be considered a likely substrate for the enzyme, and a compound with increasing levels after incubation a likely product. The identification of those compounds can therefore directly report on the enzyme's activity. This procedure should be effective for discovering novel activities of enzymes as well as identifying unknown enzymes in an unbiased manner. The results demonstrated that it was possible to monitor several known enzymatic reactions by observing the changes in substrates/products levels in a complex mixture of metabolites. This method was further applied to identify the unknown *E. coli* protein YihU as a novel hydroxybutyrate dehydrogenase, involved in an alternative succinic semialdehyde metabolism (Saito et al., 2009). In a similar approach, termed 'activity based metabolomic profiling', using LC-MS, the protein Rv1248 from *Mycobacterium tuberculosis* was identified as a 2-hydroxy-3-oxoadipate synthase (de Carvalho et al., 2010). The same group later identified the gene product of Rv1692 from *M. tuberculosis*. Originally annotated as a nucleotide phosphatase, applying the activity based metabolomic profiling method, it was shown that the enzyme is in fact a D,L-glycerol 3-phosphate phosphatase (Larrouy-Maumus et al., 2013). That this approach is also applicable outside the research of microorganisms was shown by the identification of a methyltansferase in the plant Madagascar periwinkle (*Catharanthus roseus*). The identified S-adenosylmethionine dependent N-methyltransferase plays a role in the biosynthesis of the anti cancer drug vindoline (Liscombe et al., 2010).

1.6.2. Metabolite profiling by disrupting enzyme activity

Metabolite profiling on mutants, with a gene of interest either knocked out or knocked down by RNA interference (RNAi), can also determine the function of that gene when compared to wild-type (Saghatelian et al., 2004). The most ideal case shows one metabolite increasing (the substrate of that reaction) and one metabolite decreasing (the product of that reaction). However, unless the gene of interest is essential for that organism, visible changes to the dataset can be numerous or nonexistent. Organisms could have the ability to bypass the blocked reaction and changes in the dataset can be non-conclusive. In the case of RNA interference, a complete knock-down of the gene might not be achieved and therefore not show changes in substrate and product. Compared to the *in vitro* assay (1.6.1) results from genetic mutant assays can not only give information about the gene of interest and its function, but also if that gene is essential and if loss of function causes changes to the phenotype.

1.6.3. Stable isotope labelling for pathway identification

The use of stable isotope labelling combined with untargeted metabolomics has shown to improve the interpretation of data regarding pathway identification. Normally, untargeted metabolomics lacks the information that is needed to assign detected metabolites to pathways, as metabolites can participate in multiple pathways (Fan et al., 2012). The use of labelled compounds makes it possible to follow the fate of a single metabolite within the cell. For example, by adding uniformly labelled ^{13}C glucose (U-^{13}C glucose) to the culture medium, the carbon contribution from glucose into other metabolites can be traced and its flux can also be determined (Winder et al., 2011). The uses of stable isotope labelling for novel metabolite or pathway discovery has the advantage of being independent from genetic modifications or studies with recombinant protein, therefore making it a faster approach for pathway identification (Prosser et al., 2014).

1.7. Trypanosome metabolomics

Many applications for metabolomics analysis of trypanosomes have been developed in recent years. Early trypanosome metabolomics experiments were performed with NMR, for example the assessment of the end products of anaerobic glycolysis using U-^{13}C labelled glucose (Mackenzie et al., 1983) and the tracing of polyphosphates in *T. brucei*, *T.*

cruzi and *L. major* (Moreno et al., 2000). The identification of the enzyme Acetyl:succinate CoA transferase and its role in carbohydrate metabolism in procyclics was performed using gene knock down and NMR (Rivière et al., 2004).

Earlier studies on trypanosome metabolomics was performed using direct infusion and approximately 1,000 peaks were detected, with a high percentage of peaks associated to lipid compounds (Breitling et al., 2006b).The use of LC-MS with HILIC columns optimised the identification of polar compounds in trypanosome extracts (Kamleh et al., 2008). Recently, LC-MS, using HILIC (or pHILIC) columns has emerged as the main analytical platform for trypanosome metabolomics used (in Glasgow) (Barrett et al., 2010).

Development of a minimal media (CMM) for trypanosome growth has further advanced the metabolomics platform in Glasgow, as CMM reduces background noise on the LC-MS platform used while not affecting the ability of *T. brucei's* growth (Creek et al., 2013).

For the analysis of the metabolome it is crucial to quench the parasites metabolism as rapidly as possible before extracting metabolites. The first method to achieve this was to add boiling ethanol to trypanosomes (Kamleh et al., 2008). A drawback of this method was that trypanosomes were still in media at that point, making it necessary to distinguish between intra- and extracellular metabolites. The development of new extraction methods has greatly improved the number of metabolites discovered and metabolite extraction for metabolomics is now achieved by rapidly chilling the cells in culture to 0°C by suspending them in a dry ice/ethanol bath. The medium can be removed by gently spinning the culture and washing steps before extracting the trypanosome metabolites from a cell pellet by adding a mix of chloroform:methanol:water (t' Kindt et al., 2010; Robinson et al., 2007).

Metabolomics studies lead to a better understanding of the metabolism of trypanosomes, the earlier understanding of the fatty acid synthesis metabolism could be revised and with metabolomic approaches it was shown that trypanosomes are indeed capable of fatty acid synthesis and that they use a unique mechanism to do so (Smith and Bütikofer, 2010; Stephens et al., 2007). Also, metabolomics approaches have been employed to investigate the energy metabolism in procyclic and bloodstream form trypanosomes (Coustou et al., 2008; Ebikeme et al., 2010; Mazet et al., 2013; Rivière et al., 2004). Recently it was shown that intermediates from glucose enter many branches of the metabolism, by combining stable isotope labelling with metabolite profiling (Creek et al., 2015).

Another application for the uses of metabolomics in trypanosomes are mode of action (MOA) studies for drug discovery. The recent example of Eflornithine (an inhibitor of the enzyme ornithine decarboxylase), showed that by adding Eflornithine to cultured cells and comparing treated and non treated samples, L-ornithine increased significantly in drug treated cells, while levels of putrescine decreased significantly (Vincent et al., 2012).

1.8. Secretome and Host Pathogen Interaction

1.8.1. Secretory pathway in trypanosomes

Despite trypanosomes being early divergent eukaryotes the classical secretory pathway in trypanosomes is similar to that of other eukaryotes and has been well studied for the endo- and exocytosis of VSGs. The organelles involved in the classical secretory pathway are typical eukaryotic organelles, including endoplasmic reticulum (ER), Golgi, endosomes and lysosome which are located nearby the posterior flagellar pocket (Silverman and Bangs, 2012). That trypanosomes use this secretion pathway was demonstrated for VSGs using immunogold labelling. VSGs could be traced in ER, golgi cisternae, *trans*-golgi network, transport vesicles, flagellar pocket and, ultimately, on the cell surface (Duszenko et al., 1988). The secretion of VSGs is linked to the modification of VSGs with the GPI-anchor within the ER, as deletion of the GPI anchor leads to mis-sorting into lysosomes (Triggs and Bangs, 2003).

Endocytosis is up regulated in bloodstream form trypanosomes (compared to procyclics) and important for recycling of VSGs and clearance of VSG recognising immunoglobulins (Engstler et al., 2007; Field and Carrington, 2004)(Field and Carrington, 2004). Endocytosis and exocytosis where thought to be restricted to the flagellar pocket, however, a recent study indicates that exocytosis might be possible through microvesicles (Geiger et al., 2010), a mechanism already described in *T. cruzi* (da Silveira et al., 1979).

1.8.2. Secretome

The term secretome describes a subset of the proteome, including all proteins actively secreted from the cell and the components of the cellular machinery used for protein secretion. The term (and this definition) was first used in a study that tried to predict all secreted proteins from *Bacillus subtilis* using computational methods (Tjalsma et al.,

2000). Verifying these results with proteomic approaches showed that 50% of the secreted proteins were correctly identified (Antelmann et al., 2001).

The secretome has been the focus on many studies of human pathogens to better understand the host-pathogen interaction (Ranganathan and Garg, 2009), as well as a research tool for biomarker discovery in cancer (Grønborg et al., 2006). The use of proteomic techniques for the identification of excreted/secreted proteins (ESPs) started in the1990s and became more popular from 2004 , with the first paper regarding the trypanosome excreted/secreted proteins (ESPs) or secretome published in 2008 (Holzmuller et al., 2008).

1.8.4. Host pathogen interaction

Shedding of extracellular vesicles (EVs) was demonstrated for infective trypomastigotes of *Trypanosoma cruzi*. It was further shown that shedding occurred independent from proteins present in the culture medium, but in a time and temperature dependent process (Gonçalves et al., 1991). One major component of those EVs was Tc85, a trans-sialidase/gp85 glycoprotein, which is involved in host cell adhesion and invasion (Alves and Colli, 2008). As *T. cruzi*, in contrast to *T. brucei*, has intracellular location in the mammalian host interaction with the host cell is necessary for infection.

During second stage HAT *T. brucei* crosses the blood brain barrier (BBB) in a mechanism not yet fully understood. However, secreted proteases from the bloodstream form trypanosomes, in particular cysteine proteases Cathepsin B and Cathepsin L (‘Brucipain’), play a vital role in the pathogenesis of *T. brucei*. RNAi cell lines targeting those proteases have successfully stopped the lethal infection in mice (Abdulla et al., 2008). The proteases are important for the parasites to invade the mammalian tissue after infection (Huet et al., 1992).

Secreted/excreted proteins from the parasites have been described as being important factor for virulence and to avoid the host immune response (Geiger et al., 2010). Garzon *et al*. (2006) showed that excreted/secreted proteins can inhibit the maturation of dentritic cells and stop them from inducing a lymphocytic allogenic response.

1.8.5. Trypanosome secretome

ESPs from pathogens can have a variety of functions, from protecting the cell from the host immune system, advancing the pathogenesis and to provide nutrients for cell survival. In the case of the protozoan parasite *T. brucei* the primary secretory cargo are glycosylphosphatidyl inositol anchored variable surface proteins (VSGs) which are essential for the parasite to evade the immune system (Bangs et al., 1996; Engstler et al., 2004). Several metabolic enzymes have been described as being secreted from bloodstream form trypanosomes, such as enzymes belonging to the nucleic acid metabolism, in particular IAG nucleoside hydrolase, which has been detected in bloodstream form trypanosomes (Parkin, 1996).

Bioinformatic approaches to identify ESPs exist, but are not always reliable. Proteins must have a transit peptide sequence to be secreted via the classical secretory pathway, therefore it is essential to know the secretory pathways used in a particular organism. A study on different strains of trypanosomes showed less than 20% of experimentally identified secreted proteins were predicted using bioinformatic approaches (Geiger et al., 2010). Endocytosis and exocytosis in trypanosomes occur through the flagellar pocket, but evidence has been found that trypanosomes possibly also use exocytosis of microvesicles to release proteins from cells (Geiger et al., 2010).

Proteomic approaches have been shown to give good results for the analysis of secreted proteins in leishmania and trypanosomes. Gel based proteomic analysis, like difference gel electrophoreses (DiGE), is a two-dimensional gel electrophoresis (2-DE) approach that allows for good separation of complex protein mixtures (O'Farrell, 1975). Proteins are first separated according to their charge (isoelectric focusing), followed by their size on a polyacrylamide gel. Proteins of interest can be selected from the gel, digested with trypsin and analysed using peptide mass fingerprinting. This approach, combined with fluorescent dyes is a useful tool for comparing two different sets of proteomes. A gel free approach also exists, where complex protein mixtures are digested into peptides and analysed with LC-MS/MS.

1.9. Aims

Human African trypanosomiasis is a neglected tropical disease that affects mainly the poorest regions in Africa. The disease is fatal when left untreated and drugs currently in use are old and some have undesirable side effects.

The causative agent of HAT is the parasitic protozoan *T. brucei*. The genome of this parasite has been sequenced and published. However, an estimated 40 % of the genes still have no assigned function.

This project focuses on metabolomic and proteomic approaches to determine the function of unknown metabolic enzymes and pathways in trypanosomes with three approaches being investigated:

- High throughput approach for enzyme function identification in an untargeted approach. Putatively identified metabolic enzymes from *T. brucei* strain 427, as identified from the trypanosome database TritrypDB, were used in this study.
  After the recent success stories of an unbiased, untargeted enzyme assay (Saito et al., 2009; DeCavalho et al., 2010) this approach was tested for its uses of a high throughput approach applicable to trypanosome genes.

- The use of stable isotope labelling combined with metabolomic techniques for the study of amino acid metabolism in trypanosomes, specifically the pathways involving L-methionine, L-proline and L-arginine in bloodstream form trypanosomes.

- Identification of secreted/excreted proteins from two different *T. brucei* strains, namely 427 and GVR 35, using proteomic approaches.

Chapter 2

Methods

2.1. Cell culture

2.1.1. Cell culture of *T. b. brucei*

T. b. brucei bloodstream form (strain 427) were cultured in HMI-9 medium (Gibco) (Hirumi and Hirumi, 1989) supplemented with 10% FBS Gold (PAA) or tetracycline-free FBS (Gibco), unless stated otherwise. Trypanosomes strain GVR35 were originally cultured in modified HMI-9 (from P. Voorheis; Appendix A,Table A1) supplemented with 20% FBS Gold or tetracycline-free FBS and 20% Serum plus (SAFC bioscience). For metabolomic and proteomic analyses, parasites were cultured in Creek's 'Minimal' Media, CMM ((Creek et al., 2013); Appendix A, Table A2) with 10% FBS Gold. The 2T1 cell line (Alsford and Horn, 2008) was cultured in HMI-9 medium with 10% tetracycline-free FBS and maintained in 2 µg ml $^{-1}$ puromycin and 2.5 µg ml $^{-1}$ phleomycin selection until transfection, when antibiotic selection was changed to 5 µg ml $^{-1}$ hygromycin and 2.5 µg ml $^{-1}$ phleomycin after 24 h to select for successfully transfected parasites.

For continuous growth, cell densities were kept between 5 x 10^4 and 2 x 10^6 cells ml^{-1}. The cell density was checked every 48 hours using an improved Neubauer haemocytometer. For growth curves, cells were counted every 24 hours and cell counts performed in triplicate.

Stabilates of bloodstream form trypanosomes were routinely prepared by mixing cells in mid-log phase 1:1 (v/v) with a freezing mix containing 80% culture medium and 20% glycerol. Samples were stored overnight at -80°C, before being transferred to liquid nitrogen for long term storage.

2.2. Molecular Methods

2.2.1. DNA isolation

Genomic DNA (gDNA) was extracted using the DNeasy Blood & Tissue Kit (Qiagen). Bloodstream form trypanosomes (strain 427) were grown to mid-log phase. 5 x 10^6 cells were harvested by centrifugation for 5 min at 1,800 rpm. The resulting cell pellet was resuspended in 200 µl of 1 x PBS, and DNA was extracted following manufacturer's instruction.

2.2.2. RNA isolation

RNA was extracted from bloodstream form trypanosomes using the RNeasy Mini Kit (Qiagen). Mid-log phase parasites (5 x 10^7 cells) were harvested by centrifugation for 10 min at 1,300*g*. The pellet was resuspended in 1 ml Trizol (Invitrogen) and samples stored at -80°C until RNA isolation.

For RNA isolation, to 1 ml of cells in Trizol 200 µl of chloroform was added and mixed thoroughly. After centrifuging at 12,000*g* for 15 min (4°C), the mixture separated into a lower red phenol – chloroform phase, an interphase, and an upper colourless aqueous phase. The upper colourless aqueous phase was removed and the RNeasy Mini Kit was used from this step, according to manufacturer's instruction. After RNA extraction, samples were treated with DNase (Finnzymes). RNA concentration was measured using a Nanodrop 1000 spectrophotometer (Thermo Scientific), NanoDrop 1000 software (version 3.7.0) and the method RNA-40.

2.2.3. Polymerase chain reaction

DNA amplification by polymerase chain reaction (PCR) was used to produce desired regions of DNA for molecular cloning, using appropriately designed primers, or to screen for successful transformations using vector specific primers. All oligonucleotides used where obtained from Eurofins MWG Operon (Ebersberg, Germany), and are listed in Appendix B, Table B1.

The standard PCR reaction used for colony screening was performed in 10 µl volume reactions using GoTaq DNA polymerase (Promega) and vector specific primers (T7 promoter (forward) and HIS tag (reverse)).

Cycling conditions:

1 – Polymerase activation	94°C for 5 min
2 – Denaturation	94°C for 30 s
3 – Annealing	42°C for 30 s
4 – Extension	68°C for 1 min
5 – Repeat step 2-4	30 cycles
6 – Final extension	68°C for 5 min

PCR reactions for cloning were performed in 50 µl volume using either KOD HotStart proofreading polymerase (Novagen/Merck) or Phusion HiFi proofreading polymerase (NEB). Conditions were applied as by manufacturers' instruction.

Cycling conditions (KOD HotStart proofreading polymerase):

1 – Polymerase activation	95°C for 2 min
2 – Denaturation	95°C for 20 s
3 – Annealing	54.2°C for 10 s
4 – Extension	68°C for 30 s (<1,500bp), 50 s (>1,500bp)
5 – Repeat step 2-4	30 cycles
6 – Final extension	70°C for 20 s

Cycling conditions (Phusion)

1 – Polymerase activation	98°C for 3 min
2 – Denaturation	98°C for 30 s
3 – Annealing	75°C for 30 s
4 – Extension	72°C for 30 s
5 – Repeat step 2-4	35 cycles
6 – Final extension	72°C for 10 min

2.2.4. Reverse transcription

To determine the knockdown efficiency of tetracycline induced RNAi lines, reverse transcription coupled to real-time PCR (2.2.5) was applied. For reverse transcription

reaction, 1 µg of RNA sample was mixed with 250 ng random primers (Invitrogen), 1 µl of dNTPs (10 mM) and dH_2O to a final volume of 14 µl. To remove secondary structures, the sample mix was heated to 65°C for 5 min and immediately placed on ice. To synthesise cDNA 6 µl RT-mix (containing 5x First-Strand Buffer, 0.1 M DTT and SuperScript III reverse transcriptase (Invitrogen)) was added to the sample and incubated for 25 min at 25°C, and then for a further 60 min at 50°C. The transcriptase was inactivated for 15 min at 70°C. Finally, complementary RNA was removed by adding *E. coli* RNaseH (2U) for 20 min at 37°C. Samples were prepared in duplicate, with one sample set containing dH_2O instead of SuperScript III RT as negative control.

2.2.5. Real-time PCR

Complementary DNA (cDNA) obtained from reverse transcription reaction (2.2.4) of tetracycline induced and uninduced RNAi lines were used as template DNA. SYBR® Green PCR Mastermix (Applied Bioscience) and 96-well plates were used for relative quantitative PCR. Sample setup included three replicates of each sample (+/- reverse transcription reaction) for the transcript of interest. The constitutively expressed GPI-8 transcript was chosen for comparison (Lillico et al., 2003) and water controls were used to check for contamination. Amplification of cDNA (see 2.2.4) was performed using Applied Biosystems Prism 7500 Real Time PCR system, following the 'ddCT' program for absolute quantification and marker set to SYBR. Raw data was analysed using Applied Biosystems 7500 SDS Real-Time PCR systems software.

Specific real-time PCR primers for target genes were designed using primer express software from Applied Biosystems. Primers should amplify a region of the gene in the range of 50 to 150 bp. Primer efficiency was tested and an ideal primer should have an efficiency of > 90% (Table 2.1)

Gene ID	Forward primer	Reverse primer	Primer efficiency
Tb927.7.5680	GTGTGCGTGATCGCGAAA	TGTGCCAATGCGTGATACG	98.62%
Tb427tmp.02.3040	TGTTGGAGCCGCTATTCGA	TTCACGAGAATGAAAAGCTCAAAG	96.06%
Tb427.05.3820	GCCACTGCACTGAAGGAGAAG	TGCGACCCTCAAGAAAACGT	115.08%

Table 2.1: Real-time PCR primers and their efficiency for the use in determining the knockdown effect of induced RNAi lines.

Data was obtained using Applied Biosystems 7500 Real-Time system software and relative abundance of RNA was calculated against control transcript GPI-8.

2.2.6. Plasmid generation

2.2.6.1. Plasmids for recombinant overexpression

A ligation independent cloning kit (pET 30 Xa/LIC, Novagen) was used to obtain the expression vectors for all recombinant proteins used in this study, unless stated otherwise. Constructs used are listed in Table 2.2 (a) and 2.2 (b). Primers were specifically designed for ligation independent cloning (see Appendix, Table B1) as the cloned DNA needs a vector specific overhang which can be created by treating the obtained PCR product with T4 DNA polymerase, following the manufacturer's instructions.

The annealing reaction typically contained 1 µl of Xa/LIC vector and 2 µl (0.02 pmol) of insert DNA. The reaction volume was 4 µl and included 1 µl of 25 mM EDTA.

Gene ID	Gene name	Plasmid ID
Tb427.01.1130	Glycerol-3-phosphate dehydrogenase (FAD-dependant)	pMB-G190
Tb427.06.4920	S-adenosylmethionine synthetase (putative)	pMB-G191
Tb427.08.3800	Nucleoside phosphatise (putative)	pMB-G192
Tb427.10.13130	UTP-glucose-1-phosphate uridyltransferase 2 (putative)	pMB-G193
Tb427.10.13430	Citrate synthase (putative)	pMB-G194
Tb427tmp.02.0530	Phosphoribosylpyrophosphate synthase (putative)	pMB-G195
Tb427tmp.02.3040	Aldo/keto reductase	pMB-G196
Tb427.05.3820	Aspartate carbamoyltransferase (putative)	pMB-G197
Tb427.10.2010	Hexokinase I	pMB-G198

Table 2.2 (a): **Plasmids created for recombinant protein over expression.** Proteins were used for the high throughput enzyme identification approach.

Gene ID	Gene name	Plasmid ID
Tb09.160.0810	Kynureninase (putative)	pMB-G157
Tb927.10.2750	Deoxyhypusine synthase (putative)	pMB-G158
Tb927.7.5680	Deoxyribose-phosphate aldolase (putative)	pMB-G159
Tb927.5.287b	Galactokinase, Pseudogene	pMB-G160
Tb927.2.3080	Fatty acid desaturase (putative)	pMB-G161
Tb11.01.6500	NAD+ synthase (putative)	pMB-G162
Tb927.8.2020	Arginase/agmatinase-like protein	pMB-G131

Table 2.2 (b): Plasmids created for recombinant protein over expression. Proteins were used in this study either for the high throughput enzyme identification approach or targeted investigation. Plasmids were created by B. Nijgal, with the exception of pMB-G131, which was created by E. Kerkhoven. Tb927.5.287b was annotated as Galactokinase, pseudogene. The gene sequence shows several stop codons, however, protein is produced by *E. coli* when overexpressed.

2.2.6.2. RNA interference constructs

Target sequences and oligonucleotides for cloning were designed using the program TrypanoFAN RNAit (Redmond et al., 2003) and oligonucleotides are listed in Appendix B, Table B2. Target sequence was amplified from *T. b. brucei* strain 427 gDNA using Phusion high fidelity polymerase and cloned into plasmid pGL2084 (Jones et al., 2014) in a BP recombinase (Invitrogen) reaction following manufacturer's instructions. Resulting plasmids were transformed using DH5α max efficiency cells (Invitrogen), purified and digested with *Asc*I (NEB) prior to transfection.

2.2.7. Plasmid Purification

The pET30 Xa/LIC was transformed into NovaBlue GigaSingles competent cells (Novagen) and grown on LB (Luria Broth, Sigma) agar plates supplemented with 30 µg/ml kanamycin (KAN). Up to five colonies were selected and screened by PCR for presence of the insert. A 5 ml overnight culture from a single colony, that tested positively for presence of the insert, was set up (LB medium + 30 µg/ml KAN) and the plasmid isolated using QIAprep Miniprep Kit (Qiagen GmbH, Hilden, Germany). The yield of plasmid was

assessed using Nanodrop 1000 Spectrophotometer (Thermo Scientific), using the NanoDrop 1000 software, version 3.7.0 and the method DNA-50.

2.2.8. Transformation

2.2.8.1. Competent cells

Competent cell lines were produced chemically. *E. coli* cells were streaked out on LB agar plate containing appropriate antibiotic (chloramphenicol for pLysS lines) and incubated overnight at 37°C. A single colony was inoculated in 5 ml LB broth (with appropriate antibiotic if required) and incubated overnight at 37°C. Overnight cultures were diluted 1:1000 in LB broth to a final volume of 200 ml and grown at 37°C to OD_{600} 0.6. The culture was divided into 50 ml falcon tubes, incubated on ice for 15 min, and centrifuged for 15 min, 2,000 rpm, at 4°C. The resulting cell pellets were resuspended in 16 ml RF1 buffer (Table 2.3), incubated on ice for 15 min, and centrifuged for 15 min at 1800 rpm, at 4°C. The cell pellets were pooled by resuspending in 4 ml of RF2 buffer (Table 2.3), incubated for 1 h on ice, and divided into 200 µl aliquots on dry ice. Aliquots were then stored at -80°C.

RF1 buffer	RF2 buffer
100 mM rubidium chloride	10 mM MOPS pH 6.8
50 mM $MnCl_2.4H_2O$	10 mM rubidium chloride
30 mM potassium acetate	75 mM calcium chloride
10 mM calcium chloride	15% glycerol
15% glycerol	
pH 5.8	pH 6.8

Table 2.3: RF1 and RF2 buffers for chemically competent cells

2.2.8.2. Bacterial transformation

NovaBlue GigaSingles competent cells (Merck Bioscience, UK) were used for cloning. For over-expression, several *E.coli* strains were used, depending on the protein (see 2.3.1). *E. coli* BL21 (DE3) (Merck Bioscience, UK) and Rosetta (DE3) pLysS were used for the majority of proteins.

For bacterial transformations 1-10 ng of plasmid DNA was used to transform 20 µl of competent cells. The mixture was incubated on ice for 5 min. The cells were heat shocked for 30 s at 42°C, and then chilled on ice for 2 min. A 250 µl aliquot of pre-warmed SOC medium was added to the mixture, and the cells were incubated for 60 min at 37°C with shaking. Typically, 150 µl of transformed bacteria were spread evenly over the surface of an LB agar plate containing the appropriate antibiotic, and incubated at 37°C overnight to select for transformed cells.

2.1.4. Glycerol stocks

Transformed *E. coli* stocks were prepared by adding glycerol 1:1 (v/v) to the *E. coli* culture at OD_{600} 0.6 -1.2. Glycerol stocks were stored at -80°C.

2.2.8.3. Transfection of parasites

T. brucei strain 2T1 was used for transfection (Burkard et al., 2007). Briefly, 4 x 10^7 cells from a mid-log culture were resuspended in 100 µl transfection buffer (90 mM sodium phosphate, 5 mM potassium chloride, 0.15 mM calcium chloride and 50 mM HEPES, pH 7.3 (Schumann Burkard et al., 2011)) and 10 µg linearised DNA (in sterile dH_2O) were added. Cells were electroporated in 2 mm gap cuvettes (Biorad) using the Amaxa Nucleofector II (Lonza, Germany) and the program X-001. Cells were diluted 1:20 or 1:40 in HMI-9, and seeded in 24-well plates. Appropriate antibiotics were added 12 hours post-transfection. To obtain stable clones a serial dilution on a 96-well plate was performed and clones selected and tested for the correct selection markers (Alsford and Horn, 2008).

2.3. Protein methods

2.3.1. Overexpression

For protein overexpression different strains of *E.coli* (all derived from BL21 (DE3)) were used. In general *E.coli* BL21 (DE3) and Rosetta (DE3) pLysS were used for most proteins. However, if protein overexpression failed, *E. coli* C41 (DE3) pLysS or C43 (DE3) cells were used, due to their reported ability to overexpress toxic or membrane bound proteins (Dumon-Seignovert et al., 2004). The Rosetta (DE3) pLysS, C41 (DE3) pLysS and C43 (DE3) were kindly provided by Nathaniel Jones, from the University of Glasgow. Cells were grown in LB medium at 37°C, unless stated otherwise.

Large scale overexpression was performed by inoculating 1 litre of LB (plus appropriate antibiotic) with an overnight culture, at 37°C in an orbital shaker, until an OD_{600} between 0.6 and 1.2 was reached. To induce overexpression, IPTG was added to a final concentration of 1 mM and cultures were incubated overnight at 18°C. Cells were harvested by centrifugation and the pellet was stored at -20°C if not used immediately. For small scale overexpression 70 ml of LB (plus appropriate antibiotic) were used and treated as above.

2.3.2. Protein purification

For small scale purifications, Ni-NTA Spin Columns (Qiagen GmbH, Hilden, Germany) were used, and manufacturer's instructions were followed. Large scale purifications were performed with immobilised metal affinity chromatography (IMAC) using a Poros MC20 column, stripped with 50 mM EDTA and 1 M NaCl (pH 8.0) and recharged with 0.1 M nickel sulfate prior to every run. Cells were lysed using a bacterial protein extraction solvent (B-Per, Thermo Scientific) and purified under native conditions with SB buffer (20 mM phosphate buffer plus 500 mM NaCl, pH 7.5) and increasing concentrations of imidazole (50 mM for the washing step to remove non-specific binding and 500 mM to elute the his-tagged protein of interest).

2.3.3. SDS-PAGE

Samples for SDS-PAGE were boiled in SDS buffer (50 mM Tris-HCl, pH 6.8; 2% SDS; 10% glycerol; 1% β-mercaptoethanol; 12.5 mM EDTA and 0.02% bromophenol blue) for 10 min at 95°C. Up to 20 µl protein solution was separated on NuPAGE 4-12% Bis-Tris Gel (Novex) in 1x NuPAGE MES SDS running buffer (Novex) at 125V.

2.3.4. Western blotting

Samples were run on SDS-PAGE (section 2.3.3) and transferred to a nitrocellulose membrane (Hybond-ECL) at 100 mA, for 1-2 h, at 4°C. Membranes were blocked using phosphate buffered saline/0.1% tween (PBStween) plus 5% milk powder for 1 h at room temperature, and probed with primary antibody RAD51 (1:1,000) for 1 h. Membranes were washed with PBS/tween (0.1%) three times for 10 min, incubated with secondary antibody anti-Rabbit (1:5,000) for 1 h, and washed as described. 1 ml of Piece ECL Western blotting substrate (Thermo Scientific) was added and left for 5 min before the blot was analysed.

2.3.5. Bradford assay

Protein concentrations were determined by Bradford assay using the Bio-Rad protein assay (Bio-Rad Laboratories). 50 µl of sample were mixed with 1 ml of Bradford reagent (in 1:5 dilution with dH_2O) in a 1 ml disposable cuvette. A standard curve was prepared with bovine serum albumin (BSA, Sigma-Aldrich), where 50 µl of BSA standards were also mixed with 1 ml Bradford reagent, with concentrations from 0.5 mg/ml – 0.1 mg/ml, and dH_2O as blank. The solutions were incubated for 5 min at room temperature to allow for the colour of the dye, responding to the concentration of the protein, to change. The absorbance (OD_{595}) was then determined using an Eppendorf spectrophotometer (Eppendorf, Germany). Experimental OD_{595} were plotted against concentrations of standards using an Excel spreadsheet template kindly provided by Alan Scott (Glasgow Polyomics).

2.3.6. Sample preparation for proteomics

The secretome was prepared using the method described by Holzmuller *et al*. (2008) and Grébaut *et al*. (2009). Trypanosomes were incubated at $2x10^8$ parasites/ml for 2 hours at 37°C in serum free modified HMI-9 (Voorheis, Table A1). The supernatant was separated from trypanosomes by centrifugation (1,000*g*, 10 min, 4°C) and filtered using a 0.22 µm low-binding protein filter. A mixture of protease inhibitors (Protease inhibitors complete Mini (Roche)) were added to the samples prior to storage at -80°C. Protein concentrations were measured using the Bradford assay and proteins were precipitated to concentrate the protein solutions for DiGE analysis, but also to remove protease inhibitors added to the sample sets after secretome preparation.

2.3.7. Protein precipitation

Protein precipitations for secretome sample analysis were performed by mixing the cell-free spent media with 100% acetone (1:4 v/v) and incubating overnight at -20°C, followed by 2 washes with 80% acetone at 4°C and 13,000 rpm for 10 min. Pellet was resuspended in DiGE lysis buffer (Appendix C) and stored at -80°C.

2.3.8. DiGE

Protein samples for DiGE analysis were labelled with cyanine dyes at an alkaline pH. *T. b. brucei* strain 427 was labelled with Cy5 and strain GVR 35 with Cy3. 50 µg of protein sample in 83.3 µl (GVR35) or 71.4 µl (427) DiGE lysis buffer were mixed with 400 pmol CyDye and incubated in the dark for 30 min on ice. The labelling reaction was quenched by adding 1 µl of 10 mM lysine and by incubating it for an additional 10 min. Samples were mixed, adjusted to 460 µl with DiGE rehydration buffer (Appendix C), applied to an immobilised pH gradient iso-electric focusing strip, IEF, (pH 4.0-7.0, linear range, 24cm) and run overnight according to protocol #9 (see Table 2.4).

Step 1	30V	step 'n' hold	10h
2	300V	step 'n' hold	2h
3	600V	gradient	2h
4	1000V	gradient	2h
5	8000V	gradient	2h
6	8000V	step 'n' hold	9h
7	1000V	step 'n' hold	

Table 2.4: **Protocol #9 for IEF**

Following IEF, Strips were dipped into 0.1% SDS solution and incubated in 10 ml strip equilibration buffer (SEB) + DTT (10 mg/ml) for 15 mins, and for a further 15 mins in 10 ml SEB + iodoacetamide (25 mg/ml). The equilibrated IEF strips were placed on top of a 12% SDS-polyacrylamide gel and sealed with 0.5% agarose NA in 1x running buffer (Table S5) containing a trace of bromophenol blue to act as a dye front. Gels were run at 1W per gel overnight in an Ettan DALT II system for 2-D gel electrophoresis (Amersham Bioscience). This was done with the help of Alan Scott (Glasgow Polyomics).

2.3.9. Gel imaging and image analysis

The 2-D DiGE gel was scanned using a Typhoon 9400, following manufacturer's instructions. Wavelength was set to 532 nm (Cy3) and 633 nm (Cy5) for two sequential scans and the scanner was set to high resolution scan (100 µm pixel).

Spot detection was performed using DeCycler 2D software, set for 1,000 spots recognition. Protein spots with a minimum of 2-fold difference between strain GVR 35 and strain 427

were picked for analysis. The preparatory gels were stained with Colloidal Coomassie, and images were matched with DiGE image. Protein spots of interest were manually excised from the preparatory gel in a laminar flow hood. Gel pieces were washed, first with 250 µl 100 mM ammonium bicarbonate (ABC) (twice for 30 min), followed by washes with 200 µl 50% acetonitrile (ACN) / 100 mM ABC for 45 min (wash was repeated until gel pieces were destained). 50 µl of ACN was added to gel pieces and left for 10 min before samples were dried in speed vac. Dried gel pieces were re-hydrated with 10 µl trypsin (0.02µg/µl in 25 mM ABC) and 20 µl 25 mM ABC was used to cover the gel pieces for overnight incubation. Digested proteins were dried and analysed by LC-MS/MS. Separation was achieved using an UltiMate 3000 Rapid Separation LC system (Dionex, Thermo Scientific) coupled to electrospray ionization tandem mass spectrometry (AmaZon ETD ion trap mass spectrometer; Bruker Daltonics). Raw MS/MS data were submitted to Mascot server and searched against *Trypanosoma brucei* database TriTrypDB (version 5.0).

2.3.10. Filter aided sample preparation for trypsin digest (FASP)

To analyse the whole secretome from *T. b. brucei* strain GVR 35 and strain 427, samples were prepared with FASP (Filter Aided Sample Preparation method) using an Ultracel YM-10 filter (Millipore).

20 µl of strain 427 sample (14 µg protein) and 25 µl of strain GVR 35 (15 µg protein) were used for trypsin digest. Sample volume was made up to 30 µl by adding SDT buffer (4% SDS, 100 mM Tris-HCl pH 7.6 and 0.1 M DTT). Samples were mixed with 200 µl of UA solution (8 M urea in 0.1 M Tris-HCl buffer pH 8.5), loaded into the filtration device and centrifuged for 40 min at 13,000 rpm. An additional 200 µl UA solution was added to the filter and centrifugation repeated. 100 µl of 50 mM iodoacetamide in UA solution was added to the filter and incubated at room temperature for 5 min, before being centrifuged for 30 min at 13,000 rpm. 120 µl of a solution of ammonium bicarbonate containing 0.02 µg/ml of trypsin were added to the filter and samples were incubated overnight at room temperature. Following digestion, peptides were collected by centrifugation of the filter unit for 40 min at 13,000 rpm. Samples were acidified using trifluoroacetic acid (FA), desalted and dried down for analysis in Concentrator 5301 by LC-MS/MS (UltiMate 300 Rapid Separation LC System coupled to AmaZon ETD ion trap mass spectrometer). Raw MS/MS data were submitted to Mascot server and searched against *T. brucei* database TriTrypDB (version 5.0).

2.3.11. Dimethyl-labelling for proteomics

Protein samples were treated as previously stated (2.3.10) and resulting peptides were resuspended in 100 mM triethyl ammonium bicarbonate (TEAB, Sigma-Aldrich). The GVR 35 strain sample was light-labelled with 4% formaldehyde (Sigma-Aldrich) and 600 mM sodium cyanoborohydride (Sigma-Aldrich), whereas strain 427 sample was heavy-labelled with 4% deuterated formaldehyde (Cambridge Isotope) and 600 mM sodium cyanoborohydride (Sigma-Aldrich). Samples were incubated for 1 h at room temperature after which the reaction was stopped using 1% ammonium hydroxide (Sigma-Aldrich). Labelled samples were acidified with 5% FA, dried in Concentrator 5301 and analysed by LC-MS/MS (2.3.10). The raw data was analysed with Mascot Distiller (Version 2.5.1.0). The Mascot search engine set against *T. brucei* protein database from TritrypDB. The settings were: Fixed modification was set to Carbamidomethyl (C), while Oxidation (M) was set as a variable modification. For protein identification, peptide and fragment mass tolerances were set to ± 0.3 Da allowing for two missed cleavages.

2.4. Metabolite profiling

2.4.1. *In vitro* investigation of unknown enzymes sample setup

Commercial yeast extract powder (Foremedium Ltd., Hunstanton, England, UK) was used as metabolite source. 1-2 mg of yeast extract per sample was used, and metabolites were extracted using Chloroform/Methanol/Water (ratio1:3:1 v/v/v) containing 1 µM internal standards (theophylline, 5-fluorouridine, Cl-phenyl cAMP, N-methyl glutamine, canavanine and piperazine). Samples were dried using a speed-vac and dried extracted metabolites were resuspended in 100 µl 10 mM MOPS buffer plus 5 mM $MgCl_2$ and pooled to provide an even metabolite source for control / treatment. Enzyme cofactors obtained from Sigma are listed in Table 2.5. The final concentration for each cofactor was 0.1 mM. Two sets of cofactor mixes were prepared fresh as supplements to the metabolite mix.

Cofactor 1	NAD+	NADP+	ADP	GDP	CoA	FMN	FAD+	PP	TPP
Cofactor 2	NADH	NADPH	ATP	GTP	Acetyl CoA	FMN	FAD+	PP	TPP

Table 2.5: The two sets of cofactors mixtures used for *in vitro* assay for enzyme function identification at a working concentration of 1 mM (see S1 for abbreviations).

Reactions were performed in 10 mM MOPS buffer plus 5 mM $MgCl_2$ and included the metabolite mix, cofactors for the control sample and additionally purified enzyme in the treatment sample. The mix was incubated at 37°C for 30 min and the reaction was quenched with 400 µl acetonitrile. A blank sample (10 mM MOPS buffer plus 5 mM $MgCl_2$) was similarly processed in parallel each time. Each sample included: the metabolite mix, the cofactor 1 (or 2) mix and the enzyme. Samples without the enzyme and blank samples were also prepared (Table 2.6).

	Metabolite mix	Cofactor mix 1	Cofactor mix 2	Enzyme	ACN	Enzyme
Sample 1 -E	86 µl	10 µl		4 µl	400 µl	
Sample 1 - C	86 µl	10 µl			400 µl	4 µl
Sample 2 -E	86 µl		10 µl	4 µl	400 µl	
Sample 2 - C	86 µl		10 µl		400 µl	4 µl
Blank - Mops					400 µl	

Table 2.6: Sample preparation for *in vitro* investigation.

2.4.2. Intracellular metabolite extraction from parasites

Trypanosomes were grown to mid-log phase and a sample volume equivalent to 5 x 10^7 cells was rapidly cooled to 4°C by submerging the 50 ml falcon tube into a dry ice/ethanol bath. Samples were kept on ice (or 4°C) from this step onwards. Samples were centrifuged at 1,250*g* for 10 min, and most supernatant, except 1 ml, was removed. The pellet was resuspended in the remaining 1 ml of medium. This was transferred to an Eppendorf tube and briefly centrifuged to completely remove the supernatant. The cell pellet was washed in cold 1xPBS and metabolites extracted by resuspending in 100 µl chloroform:methanol:water (ratio1:3:1 v/v/v) with internal standards (2.4.1) and by vigorously shaking for 1 h at 4°C. Samples were centrifuged at 16,000*g* for 10 min and the supernatant collected and stored at -80°C under argon.

2.4.3. ^{13}C – labelled tracking

T. b. brucei strain 427 was grown in CMM + 10% FBS Gold (PAA, Piscataway, NJ) for U-^{13}C-labelled tracking. For L-methionine studies, 50% of U-^{13}C L-methionine (50 µM L-methionine and 50 µM U-^{13}C L-methionine) was added to the parasite culture medium (starting density of 2 x 10^4 cells ml^{-1}). For L-proline and L-arginine, 100% labelled

compound was used at a concentration of 200 µM. Cells were incubated at 37°C for 48 h, and metabolites extracted as previously described (2.4.2). Samples were prepared in triplicate and for every labelled setup a control with the equal concentration of unlabelled compound was set up, also in triplicate. Fresh medium and spent medium controls were also collected and 5 µl were added to 100 µl extraction solvent (CMW 1:3:1) prior to analysis.

Labelled compounds were obtained from Cambridge Isotope Laboratory, Inc.:

L-methionine, $^{13}C5$, enrichment 99%, cat: CLM-893-H-0.1

L-proline, $^{13}C5$, enrichment 99%, cat: CLM-2260-H

L-arginine, $^{13}C6$, enrichment 99%, cat: CLM-2265-H-0.1

2.4.4. Standards

Internal standards were added to the extraction solvent as previously described (2.4.1): during the analysis the stability of the samples can be tracked by comparing the total ion current (TIC) profile of samples and internal standard. With each experiment a set of authentic standards was run prior to the sample set (Appendix D, Table D1). The authentic standards are used for metabolite identification and retention time prediction in IDEOM as described by Creek *et al*. (2011).

2.4.5. Liquid chromatography-mass spectrometry

The sample platform chosen for this project was liquid chromatography coupled with mass spectrometry. All samples were separated with high performance liquid chromatography (HPLC) on either ZIC-HILIC (Hydrophilic Interaction Liquid Chromatography column, Merck) or ZIC-pHILIC (polymer based - Hydrophilic Interaction Liquid Chromatography column, Merck) prior to mass detection on an Exactive Orbitrap mass spectrometer (Thermo Fisher). Analysis was performed in positive and negative mode, using 10 µl injection volume and a flow rate of 100 µl/min. For HPLC gradient, ZIC-HILIC solvent A was 0.1% formic acid in water and solvent B was 0.08% formic acid in acetonitrile. ZIC-pHILIC solvent A was 20 mM ammonium carbonate in H_2O, and solvent B was 100% acetonitrile.

2.4.6. Data analysis

The analysis was performed using IDEOM software (Creek et al., 2012b). IDEOM uses the software MZmatch (Scheltema et al., 2011) and the statistical platform R to analyse the raw data. First, the raw data is converted to mzXML files using the software msconvert. The program XCMS can use the mzXML files to identify peaks and convert the data into peakML files. MZmatch uses peakML files to group peaks across replicate samples, filter peaks by removing groups with high variability in peak intensity, and annotate related peaks. Those MZmatch files will be used by IDEOM to run an automated identification procedure based on exact mass and retention time. Furthermore, IDEOM assigns confidence levels to the identifications, relying on authentic standards run within each experiment.

mzXML files can also be used for analysis of metabolite tracking with uniformly ^{13}C-labelled compounds. The open source software mzMatchISO (Chokkathukalam et al., 2013) works within MZmatch.R. The labelled compounds have the same retention time as their unlabelled counterparts, but differ in the mass of the heavy/light carbons when labelling occurs. To assign the detected mass shift, it requires a tab-delimited text input file containing the compounds of interest to be matched against the dataset, to search for the labelled compounds and their labelling pattern. The output file is in pdf format, showing the detected labelling pattern and peak shapes.

2.5. Enzyme assays

2.5.1. Hexokinase

Hexokinase (Sigma) activity was assayed in a coupled reaction with glucose-6-phosphate dehydrogenase (Figure 2.1). The reaction mix included 50 mM Tris-HCl, pH 8.0, 13.3 mM magnesium chloride, 670 mM glucose, 16.5 mM ATP, 6.8 mM $NADP^+$ and 1 IU glucose-6-phosphate dehydrogenase. The reaction mix was run at 25°C for 6 min on a UV-VIS spectrophotometer (UV-2550, Shimadzu) before hexokinase was added and changes in absorbance (A_{340}) measured over time.

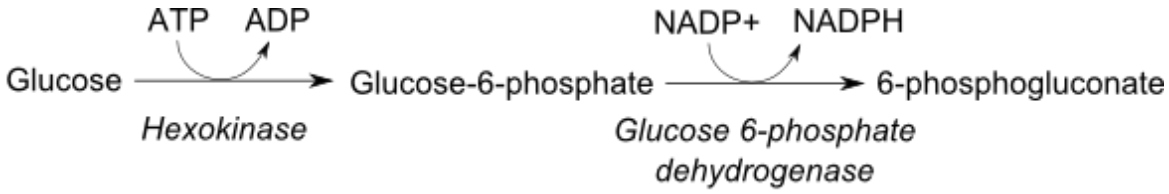


Figure 2.1: Coupled enzyme assay for hexokinase. The phosphorylation of D-glucose was measured by production of NADPH at 340 nm.

2.5.2. Glucose dehydrogenase

Glucose dehydrogenase (Figure 2.2) was ordered from Sigma and made up to 1 U/mg with 10 mM Tris-HCl buffer (pH 7.2). Activity was tested using a UV-VIS spectrophotometer (UV-2550, Shimadzu) the reaction mix included: 100 mM Tris-HCl buffer, 1 M glucose and 20 mM NAD^+.

Reaction mix was run on the spectrophotometer at 25°C for 5 min before glucose dehydrogenase was added. Changes in absorbance (A_{340}) were measured over time.

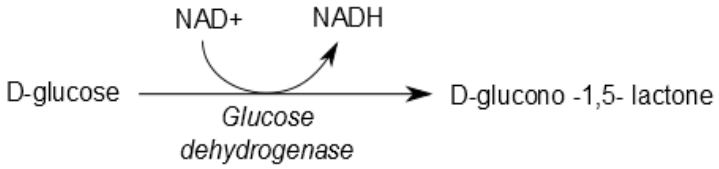


Figure 2.2: Reaction catalysed by glucose dehydrogenase. The formation of NADH was measured at 340 nm.

2.5.3. NAD^+ synthase

NAD^+ synthase activity was assayed in a coupled reaction assay (Figure 2.3). In the first reaction NAD^+ was synthesised by either (1) reaction mix containing 2 mM ATP, 5 mM magnesium chloride, 50 mM Tris-HCl, 56 mM potassium chloride, 1 mM deamido-NAD^+, 20 mM L-glutamine and 0.2 mg/ml BSA, pH 8.0 (Wojcik et al., 2006), or (2) reaction mix including 2 mM ATP, 2 mM ammonium chloride, 20 mM magnesium chloride, 2 mM potassium chloride, 1 mM deamido-NAD^+ and 0.2 mg/ml BSA (Ozment et al., 1999).

In both cases the reaction volume was 100 µl.

Reaction mix was incubated for 1 h at 37°C with recombinant protein and the reaction stopped by boiling the mix for 3 min and immediately cooling it down on ice. Reactions were spun down for 10 min at 12,000 rpm (4°C) and 90 µl were used for further analyses.

To determine if the putative NAD^+ synthase had produced NAD^+, 90 µl of reaction mix obtained from the first assay were used instead of NAD^+ in a glucose dehydrogenase assay (2.5.2).

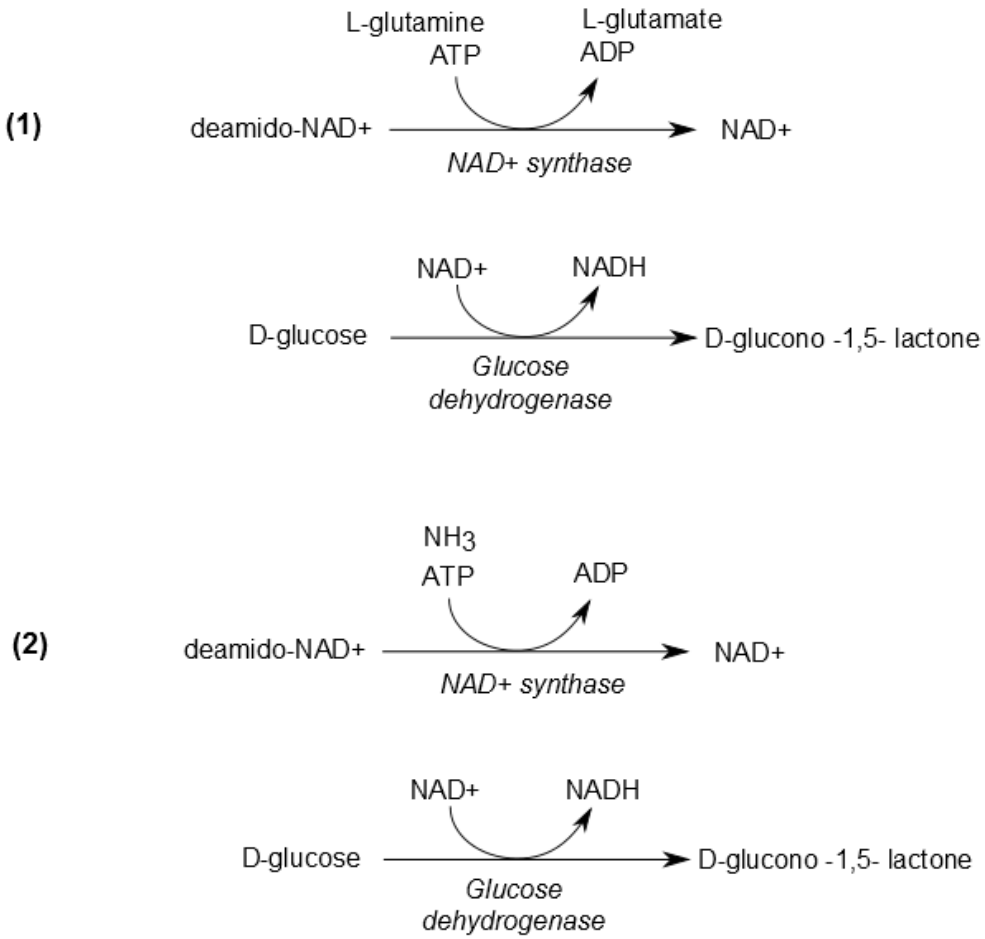


Figure 2.3: Coupled enzyme assays for NAD^+ synthase. (1) Glutamine dependend NAD^+ synthase assay (Wojcik et al., 2006) and (2) Ammonia dependent NAD^+ synthase assay (Ozment et al., 1999). In both assays the formation of NADH from NAD^+ was measured at 340nm**.**

Chapter 3

3.1. Introduction

In this chapter the use of metabolomics for enzyme function identification will be explored and discussed. Although bioinformatics studies can help to identify the function of a protein, using tools such as BLAST alignments and domain searches, having experimental proof is required and important when studying the metabolism of any organism.

Enzymes used for this study were chosen randomly from a list of putative enzyme that was obtained from the trypanosome database TritrypDB. The only objective for the chosen 'enzymes' was that they should have a predicted metabolic function, relatively small size, so the cloning and over-expression would not cause too many difficulties, and that their function had not been determined before using recombinant or purified protein. The goal was to set up a workflow that would allow a high throughput approach to screen putative trypanosome proteins for enzymatic reactions. Certain compromises had to be made to make this approach fast with relatively low costs. The choice for a good over expression system was one of the compromises. *E. coli* is a well established organism to over express proteins; however, the lack of post translational modifications could cause problems for the functionality of the enzymes. Other over expression systems, for example trypanosomatid expression systems (Tetaud et al., 2002), would therefore be more suitable. However, high cost of culture media and lower levels of over expression was seen as too big a disadvantage compared to *E. coli*. Several *E. coli* strains were tested in this study. Originally, BL21 (DE3) were used (E. Kerkhoven, thesis), but Rosetta (DE3) cells showed higher yield and allowed for a higher number of expressed proteins and from 21 proteins originally selected for the screen, seven could be analysed.

The metabolite basis of this assay was a commercially available yeast extract. Trypanosome extract would have been preferred, but as with the expression system, the cost of media and the low density of cells in culture (compared to yeast or *E. coli*) made other metabolite sources better suited. *E. coli* extract, marmite and several stock cubes were also tested for their use, but the commercial yeast extract seemed the better choice for this screening method.

(1) Yeast extract was commercially obtained at a high quantity, meaning the standard condition was highly controlled for all experiments.

(2) Yeast extract was treated to remove all proteins; therefore it was certain that the only protein added to that mix was the intended purified enzyme.
(3) Yeast extract provided a high number of detected metabolites on both ZIC-HILIC and ZIC-pHILIC while being fast and easily prepared (compared to *E. coli* extract).

3.1.1. *In vitro* assay combined with metabolite profiling

Using the screening method developed by Saito *et al.* (2006), seven (out of 21) putative identified enzymes from *T.b.brucei* (see Table 3.1) were tested for their function using metabolite profiling by Liquid Chromatography – Mass Spectrometry (LC-MS). The enzymatic activity is determined by monitoring the changes in metabolite levels between control (no enzyme) and treatment (incubated with enzyme) samples.

Yeast extract was used as the metabolite source and to ensure that possible essential cofactors needed for enzymatic reactions were present, mixes of most common cofactors (as listed in method 2.4.1) were added to every reaction mix. As described, two cofactor mixes were prepared and samples were set up in two batches each containing a different cofactor mix.

Gene ID	Gene name	Plasmid ID
Tb927.10.2750	Deoxhypusine synthase, putative	pMB-G158
Tb927.7.5680	Deoxyribose-phosphate aldolase, putative	pMB-G159
Tb11.01.6500	NAD synthase, putative	pMB-G162
Tb427.06.4920	S-adenosylmethionine synthetase, putative (METK1)	pMB-G191
Tb427.10.13430	citrate synthase, putative	pMB-G194
Tb427tmp.02.3040	aldo/keto reductase, putative	pMB-G196
Tb427.05.3820	aspartate carbamoyltransferase, putative	pMB-G197

Table 3.1: Identification of putative identified enzymes used in high throughput approach

Ideally, decreasing level of a specific metabolite indicates the substrate of an enzymatic reaction, while an increasing level indicates the product of that reaction. However, due to the ability of mass spectrometry to simultaneously identify a large number of metabolites (the LC-MS approach of this study routinely detects around one thousand metabolites per experiment), changes in several, unrelated, metabolites will be detected. This is due to sample variation between either a set of controls or even between conditions. To keep those changes to a minimum, samples were run in triplicate.

To validate the enzyme assay two commercially available enzymes were tested using this approach. The enzymes chosen were glucose dehydrogenase and hexokinase (both Sigma).

3.1.2. RNA interference

To further assess potential function of the proteins, when the *in vitro* assay approached failed, and to establish if the protein is essential to bloodstream form trypanosomes, RNA interference (RNAi) lines were created. *T. brucei* strain 2T1 were transfected with plasmid pGL2084 (Jones et al., 2014) containing a fraction of the gene of interest in a stem loop construct (Alsford and Horn, 2008) and RNAi was induced with either 1 µg ml^{-1} or 5 µg ml^{-1} tetracycline. To determine whether protein was essential or not, growth curves were made by counting uninduced and induced cells every 24 h. To establish a possible function of the protein, metabolite profiling on extracted metabolites from induced cultures were compared to the metabolite set of uninduced cultures.

RNAi lines were created for four putative identified enzymes: deoxyribose-phosphate aldolase, NAD^+ synthase, also alpha/keto reductase and aspartate carbamoyltransferase. With the exception of NAD^+ synthase, which did not obtain puromycin sensitivity and was therefore not further analysed, all cell lines showed the correct selection markers and knock down effect was assessed using reverse transcription RT-PCR.

Using metabolite profiling on RNAi lines, ideally, a decreasing metabolite (comparing induced cell line to uninduced cell line) should indicate the product of an enzymatic reaction, while an increasing metabolite should indicate the substrate of that reaction. However, as the whole cell metabolome is analysed, blocking one enzyme in the system could have a knock on effect on other reactions, therefore making it difficult to see the primary reaction.

3.2. Results

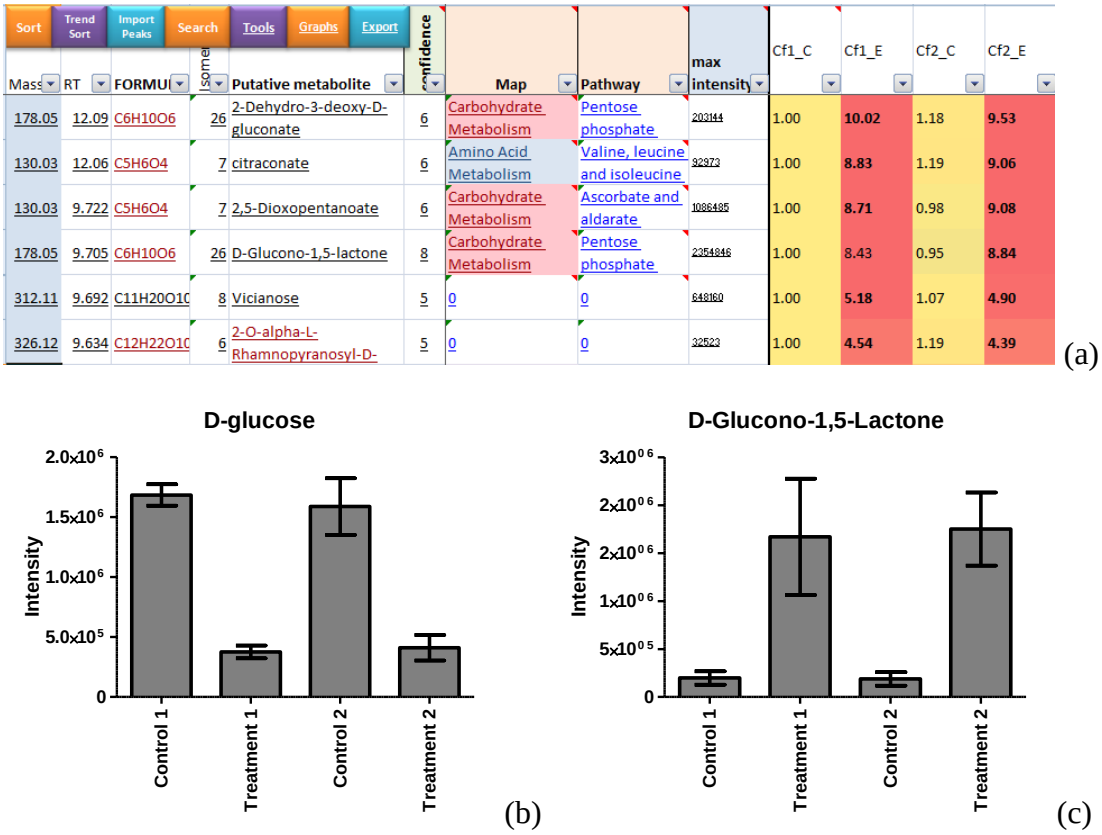
3.2.1. Validation of *in vitro* method

3.2.1.1. *In vitro* investigation of glucose dehydrogenase by metabolite profiling

Glucose dehydrogenase is an enzyme belonging to the family of oxidoreductases and acts on the CH-OH group of the donor molecule. It catalyses the following reaction:
Glucose + acceptor ⟷ D-Glucono-1,5- lactone + reduced acceptor
The acceptor can be FAD, NAD(P) or pyrroloquinoline quinine (PQQ) and depending on the cofactor use the enzyme is categorized into EC group 1.1.1, 1.1.99 or 1.1.5 respectively (Ferri *et al*., 2011). Glucose dehydrogenase used here was obtained from Sigma and was isolated from *Pseudomonas sp*. with the EC 1.1.1.47, suggesting NAD^+ and $NADP^+$ the cofactors used by this enzyme according to BRENDA.

Data analysed using IDEOM showed five metabolites significantly increased (Figure 2.1, a) in both samplesets. This would suggest either that $NAD(P)^+$ was either already present in the yeast extract or the use of a different cofactor present in both sampleset (for example FAD). From the putatatively identified metabolites, only D-glucono-1,5-lactone was a basepeak. The related peak function from mzMatch assigned the putatatively identified metabolites 2,5-Dioxopentanoate, Vicianose and 2-O-alpha-L-Rhamnopyranosyl-D-glucopyranose to the D-glucono-1,5-lactone basepeak. The very similar retention times and of all four metabolites makes it likely that the changes are related or even that the peaks resemble the same metabolite, with differences in mass caused by fragmentation or complex formation. 2,5-Dioxopentanoate seems to be a fragment of D-glucono-1,5-lactone as the peak intensity pattern of every individual replicate is identical. Furthermore, the two metabolites 2-Dehydro-3-deoxy-D-gluconate and citraconate show the same pattern. Interestingly, 2-Dehydro-3-deoxy-D-gluconate has the same mass as D-glucono-1,5-lactone, as have citraconate and 2,5-Dioxopentanoate. However, the retention times are different. A possible explanation for this is that D-glucono-1,5-lactone comes of the column at two different times, as sometimes seen with D-glucose. However, as D-glucono-1,5-lactone is not in the authentic standard list, there is no proof for this. Only one metabolite was decreased significantly in this dataset. It was putatatively identified as 2-C-Methyl-D-erythritol 4-phosphate. However, closer investigation revealed that this is most likely an adduct of D-glucose, as the retention time of 14.65 min matches the authentic standard.

As this assay was performed with a known enzyme, the decrease in glucose levels (Figure 3.1, b) and the increase of D-glucono-1,5-lactone (Figure 3.1., c) can be seen as confirmation for the validation of this assay. However, it also highlights the problematic of using complex datasets, such as metabolomics, to identify enzyme functions.



Mass	RT	FORMULA	Isomers	Putative metabolite	Confidence	Map	Pathway	max intensity	Cf1_C	Cf1_E	Cf2_C	Cf2_E
178.05	12.09	C6H10O6	26	2-Dehydro-3-deoxy-D-gluconate	6	Carbohydrate Metabolism	Pentose phosphate	203144	1.00	10.02	1.18	9.53
130.03	12.06	C5H6O4	7	citraconate	6	Amino Acid Metabolism	Valine, leucine and isoleucine	32973	1.00	8.83	1.19	9.06
130.03	9.722	C5H6O4	7	2,5-Dioxopentanoate	6	Carbohydrate Metabolism	Ascorbate and aldarate	1086485	1.00	8.71	0.98	9.08
178.05	9.705	C6H10O6	26	D-Glucono-1,5-lactone	8	Carbohydrate Metabolism	Pentose phosphate	2354846	1.00	8.43	0.95	8.84
312.11	9.692	C11H20O10	8	Vicianose	5	0	0	648160	1.00	5.18	1.07	4.90
326.12	9.634	C12H22O10	6	2-O-alpha-L-Rhamnopyranosyl-D-	5	0	0	32523	1.00	4.54	1.19	4.39

(a)

Figure 3.1: Significant changes in the metabolomics dataset of glucose dehydrogenase assay. Assay was set up as described in 2.4.1. and analysed with IDEOM. (a) Screenshot of IDEOM spreadsheet showing the most significantly increased metabolites in the dataset (highlighted in red) compared to a no enzyme control. (b) peak intensity of the substrate and (c) peak intensity of the product of glucose dehydrogenase. Abbreviations: Cf1_C: Cofactor 1, Control (no enzyme); Cf1_E: Cofactor 1, Treatment (with enzyme); Cf2_C: Cofactor 2, Control and Cf2_E: Cofactor 2, Treatment.

3.2.1.2. *In vitro* investigation of hexokinase by metabolite profiling

Hexokinase is the starting enzyme of the glycolysis and catalyses the following reaction:

Glucose + ATP ↔ Glucose 6-phosphate + ADP

Previous datasets on commercial enzymes (data not shown) showed many changes, specifically many metabolites decreasing in the treatment samples (most of them peptides), making an analysis difficult. The enzymes used for those assays had been already in use in our lab for a while, so enzymes were ordered fresh and assays were repeated. However, due to the changes mentioned above observed only in enzyme treated samplesets, a second extraction step was added to the workflow to test if those changes might disappear. For the hexokinase assay results shown here, four replicates for each condition were prepared and

two of the replicates were vortexted in the cold room (4°C) for 30 min after the enzymatic reaction was quenched with acetonitrile (ACN). Data analysis comparing those two conditions showed overall fewer 'missing' metabolites in both conditions. However, peak intensities were increased in samples with the added extraction step.

Unlike the results of glucose dehydrogenase, as shown above, this dataset did not show the distinction in changes between the product and the substrate, which could be attributed to the high levels glucose in the sample mix and the fact that ATP was not detected. As ATP was added to the sample set 2, the explanation for no peak being detected in IDEOM might be that due to low intensity that peak was filtered out by IDEOM. However the product of the reaction, glucose 6-phosphate is only detected in enzyme treated cofactor mix 2 (Figure 3.2 (b) and (d)). Also important, the enzyme treated cofactor 2 sample shows high levels of ADP, which is not present in cofactor 2 control sample (Figure 3.2 (a) and (c)). These results conclude that the added enzyme is indeed a hexokinase, as both products can be detected in the dataset.

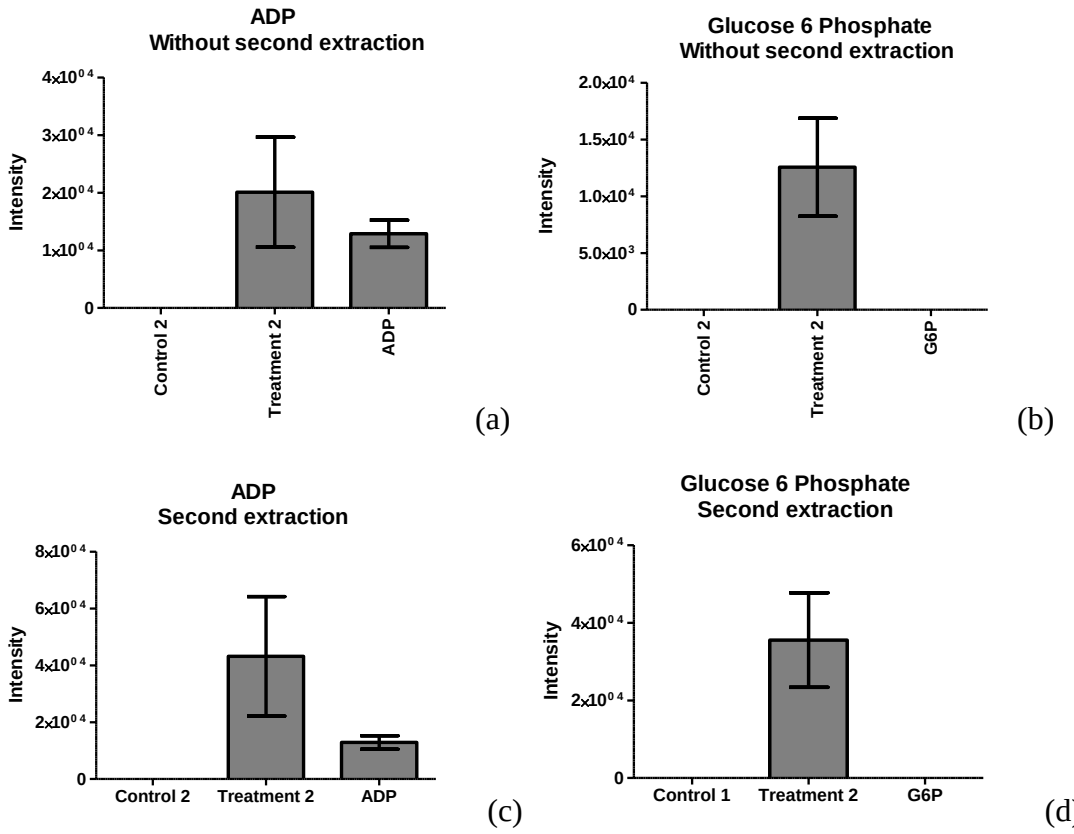


Figure 3.2: Significant changes in the metabolomics dataset of hexokinase assay. Assay was set up as described in 2.4.1. and analysed with IDEOM. Most significant changes in the hexokinase dataset were ADP and glucose-6-phosphate (G6P) increased in the sample set 2. Sample set 2 had ATP added, which is required for hexokinase activity. Metabolite levels shown on the right (ADP or G6P) are the peak intensities from the control sample without ATP. (a) and (b) shows peak intensities of compounds without second extraction step, while (c) and (d) were incubated for 30 min at 4°C after enzymatic reaction was quenched.

3.2.2. Cofactor stability

As cofactors were not consistently detected using IDEOM for data analysis and cofactors present in the authentic standard mixes reliably failed to be detected, a study was done to investigate the stability of the cofactor stock solutions (stored at -20°C). The two cofactor mixes were made fresh and an aliquot was taken for metabolite extraction and immediate analysis by LC-MS. The remaining samples were stored at -20°C and reanalysed every week for 4 weeks.

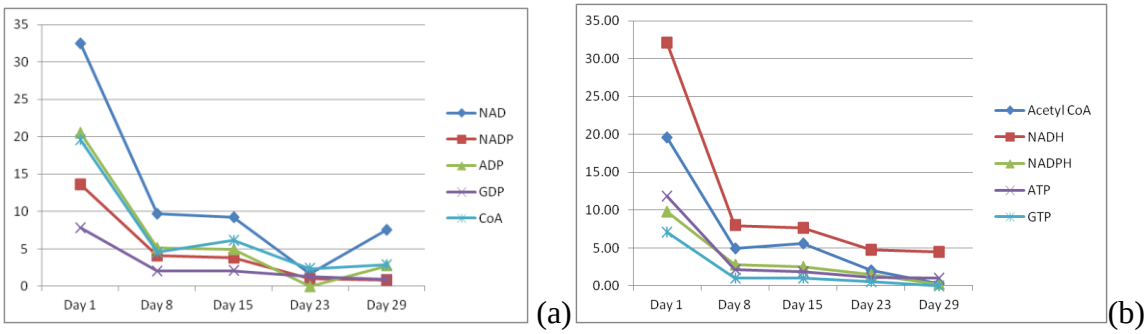


Figure 3.3: Decreasing levels of detected cofactors from cofactor mix 1 (a) and 2 (b) over a time course of four weeks. Samples were tested fresh at Day 1 and re-tested after 8,15,23 and 29 days. Analysis was performed by normalising the identified metabolite against an internal standard (N-methyl glutamine) using IDEOM.

Figure 3.3 shows the detected levels of the cofactors decreasing dramatically from day 1 to day 8. If the cofactor mixtures degrade at -20°C or if the freeze/thaw cycles have an impact on the compounds being detected is not clear. However, this clearly indicates that cofactor mixes must be freshly prepared for the use in the *in vitro* enzyme assay. The stability test shows that the cofactors used can be detected (except for PP and TPP) on the LC-MS platform using ZIC-pHILIC columns. However, why are they not as easily detected in the IDEOM spreadsheets for the *in vitro* assays?

One possible explanation is that the cofactor mixes used only contain 1mM of each cofactor, while the cofactor stability used the stock solution of 10mM. At the end of each enzyme assay incubation the sample was further diluted which would make the cofactor concentration in the analysed sample even lower. Figure 3.4 shows the differences in peak intensities of ATP, when run at 10mM and 1mM. It also seems plausible that in complex mixtures IDEOM filters the cofactor peaks out as they might appear under the threshold limit.

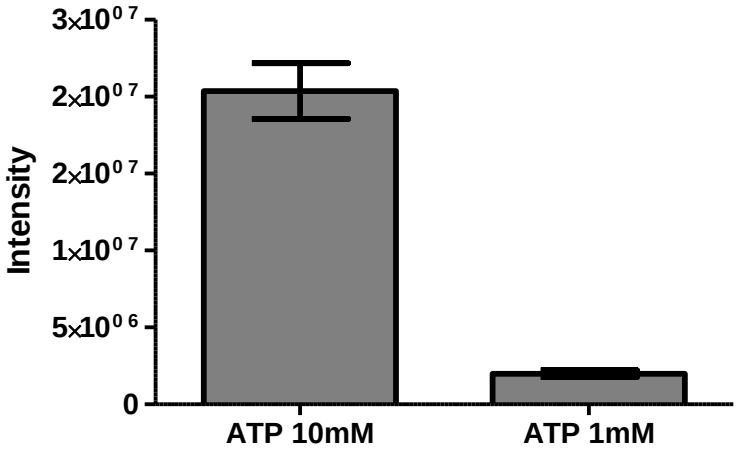


Figure 3.4: Peak intensities of ATP, 10mM and 1mM at day 1 of the time course.

The cofactors, as listed in Methods section (or Table 3.2), are present in the standard mixes that are analysed with every MS run. However, in complex mixtures they are only rarely detected and it would be beneficial to have an idea of the retention time and also how reliable the detection of these compounds is. Table 3.2 shows the mass and retention time detected for each cofactor over a time course of four weeks. As this experiment was targeted (only one cofactor per sample) and as mentioned above cofactors in the standards are only rarely detected the confidence of the peak identification was judged by detected mass, retention time similar to the expected one and peak intensities. ATP and ADP were consistently detected over 4 weeks, but also traces of ADP and AMP were detected in the ATP sample, and ATP and AMP in the ADP sample. Although these findings are not necessarily surprising it is important to keep in mind when using different cofactor mixes for enzyme assays. GDP, GTP, NAD and NADH were consistently detected, except for the last time point where GTP and NAD were not. NADP could not be detected in an early time point in the IDEOM spreadsheet (Day 7), which was surprising as the following time points showed clear levels of NADP in the sample. Coenzyme A, acetyl-coenzyme A, FAD and FMN were consistent in mass and retention time. Only PP and TPP were inconsistently detected and those compounds might be not suitable for pHILIC separation.

	1		2		3		4		5	
	Mass	RT	Mass	RT	Mass	RT	Mass	RT	Mass	RT
ATP	506.995	18.949	506.995	18.933	506.995	18.792	506.995	18.971	506.995	19.268
ADP	427.03	17.388	427.03	17.394	427.03	17.34	427.03	17.787	427.03	18.065
GTP	522.99	21.843	522.99	21.851	522.99	21.669	522.99	21.949		
GDP	443.024	20.205	443.024	20.269	443.024	20.188	443.024	20.531	443.024	20.78
NAD	663.109	16.076	663.109	16.108	663.109	15.92	663.109	15.98		
NADH	665.125	15.316	665.125	15.263	665.125	15.169	665.125	15.24	665.125	15.43
NADP	743.075	18.772			743.075	18.673	743.075	18.748	743.075	18.914
NADPH	745.091	19.079	372.545	19.093	372.546	18.922	372.545	18.954	745.091	19.195
CoA	767.115	15.544	383.558	15.595	767.115	15.357	767.115	15.482	767.115	15.778
Acetyl-CoA	809.126	14.226	809.126	14.233	809.126	14.014	809.126	14.09	809.126	14.275
FAD	785.157	13.159	785.157	13.217	785.157	12.975	785.157	12.978	785.157	13.109
FMN	456.105	13.128	456.105	13.167	456.105	12.941	456.105	12.962	456.105	13.121
PP	247.025	21.124			247.025	21.288				
TPP					424.037	17.102	424.037	17.449		

Table 3.2: Detected mass and retention time from each cofactor during a timecourse of four weeks. Timepoint 1: Day 1 (when cofactor stocks were prepared), Timepoint 2: Day 8, Timepoint 3: Day 15, Timepoint 4: Day 23, Timepoint 5: Day 29.

	Monoisotopic molecular weight
Acetyl-CoA	809.126
ADP	427.029
ATP	506.996
CoA	767.115
FAD	785.157
FMN	456.105
GDP	443.024
GTP	522.991
NAD	664.117
NADH	665.125
NADP	744.083
NADPH	745.091
PP	247.025
TPP	425.045

Table 3.3: Monoisotopic molecular weight of the cofactors used in this study, taken from metacyc.org

3.2.3. Results putative enzymes

From 21 genes originally selected from TritrypDB, 16 were successfully amplified using PCR. Those PCR products were treated with T4 DNA polymerase and annealed into pET 30 Xa/LIC over expression vector, except for gene Tb927.8.2020, which was cloned into pET28 vector (Kerkhoven, PhD thesis, 2012). Five PCR products did not get the DNA amount needed for T4 DNA polymerase treatment and these were not included in further experiments (for all gene ID, see Table 3.4).

Plasmid DNA of the cloned constructs were transformed into *E. coli* BL21(DE3) for overexpression, followed by small scale protein purification using nickel affinity columns (Ni-NTA Spin Columns (Qiagen)) for initial over-expression screen.

Clones pMB-G161, pMB-G192, pMB-pMB-G195 did not over-express in BL21(DE3) or Rosetta (DE3) pLysS.

Over-expression of clones pMB-G157, pMB-G190, pMB-G193 and pMB-G198 was achieved in Rosetta (DE3) pLysS, however, protein was not detected in soluble fraction and therefore those proteins were not used for further analyses.

Clone pMB-G160 is unnotated as a pseudo-gene in TriTrypDB, meaning that additional stop codons were found. Protein over-expression showed a much smaller protein than originally expected and this was not analysed further. For the nine proteins remaining, higher yield was achieved in *E. coli* Rosetta pLysS and were therefore used for the *in vitro* assay (Figure 3.5)

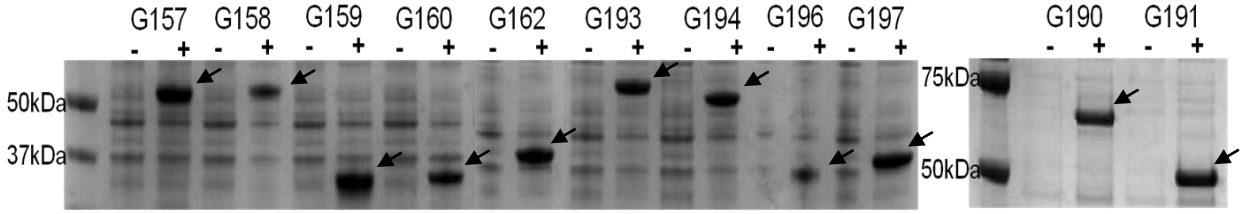


Figure 3.5: Whole cell extract of Rosetta (DE3) pLysS pre (-) and post (+) induction with IPTG (1mM final). Expected recombinant protein sizes were G157: 53kDa, G158: 55kDa, G159: 35kDa, G160: 55kDa (Pseudogene), G162: 38kDa, G190: 71kDa, G191: 49kDa, G193: 59kDa, G194: 59kDa, G196: 40kDa, G197: 41kDa and overexpressed proteins are indicated with an arrow.

Originally, over expressed proteins were purified using Ni-NTA Spin Columns (Qiagen). However, the resulting eluates showed high contamination with other proteins. Proteomic analysis of one set of purified proteins contained from the spin columns even showed enzymes from *E. coli* and changes to the metabolomics dataset could be traced to one of them. Therefore, protein purification was changed to use HPLC with a Poros MC20 column (2.3.2).

Gene ID	Gene name (putative)	Plasmid ID	PCR for cloning	Protein over expression	Metabolomics assay
Tb927.8.2020	Arginase/agmatinase-like protein	pMB-G131	E. Kerkhoven, PhD thesis, 2012	High levels overexpression	see 3.2.3.8
Tb09.160.0810	Kynureninase	pMB-G157	performed by B. Nijgal	Not soluable	n.a.
Tb927.10.2750	Deoxyhypusine synthase	pMB-G158	performed by B. Nijgal	High levels overexpression	see 3.2.3.3
Tb927.7.5680	Deoxyribose-phosphate aldolase	pMB-G159	performed by B. Nijgal	High levels overexpression	see 3.2.3.4
Tb927.5.287b	Galactokinase, Pseudogene	pMB-G160	performed by B. Nijgal	smaller size than expected	n.a.
Tb927.2.3080	Fatty acid desaturase	pMB-G161	performed by B. Nijgal	No over expression	n.a.
Tb11.01.6500	NAD^+ synthase	pMB-G162	performed by B. Nijgal	High levels overexpression	see 3.2.3.5
Tb427.01.1130	Glycerol-3-phosphate dehydrogenase	pMB-G190	0.2 pmol	No over expression	n.a.
Tb427.06.4920	S-adenosylmethionine synthetase	pMB-G191	0.2 pmol	High levels overexpression	see 3.2.3.1
Tb427.07.4390	Threonine synthase	n.a.	No positive PCR	n.a.	n.a.
Tb427.08.3800	Nucleoside phosphatase	pMB-G192	0.2 pmol	No over expression	n.a.
Tb427.10.12980	Methyltransferase	n.a.	No positive PCR	n.a.	n.a.
Tb427.10.13130	UTP-glucose-1-phosphate uridylyltransferase	pMB-G193	0.2 pmol	Not soluable	n.a.
Tb427.10.12430	Citrate synthase	pMB-G194	0.2 pmol	High levels overexpression	see 3.2.3.2
Tb427tmp.01.3640	Acyl-CoA dehydrogenase	n.a.	No positive PCR	n.a.	n.a.
Tb427tmp.02.0530	phosphoribosylpyrophosphate synthetase	pMB-G195	0.2 pmol	Not soluable	n.a.
Tb427tmp.02.3040	Aldo/keto reductase	pMB-G196	0.2 pmol	High levels overexpression	see 3.2.3.6
Tb427.10.2490	Glucose-6-phosphate 1-dehydrogenase	n.a.	No positive PCR	n.a.	n.a.
Tb427.05.3820	Aspartate carbamoyltransferase	pMB-G197	0.2 pmol	High levels overexpression	see 3.2.3.7

Gene ID	Gene name (putative)	Plasmid ID	PCR for cloning	Protein over expression	Metabolomics assay
Tb427.10.2010	Hexokinase	pMB-G198	0.2 pmol	Not soluable	n.a.
Tb427.05.4560	Guanine deaminase	n.a.	No positive PCR	n.a.	n.a.

Table 3.4: Trypanosome genes selected for *in vitro* enzyme identification assay. Gene ID represents the identification number given by TritrypDB. All genes chosen represented putative identified enzymes, except for glycerol-3-phosphate dehydrogenase, glucose-6-phosphate 1-dehydrogenase and hexokinase which were already descript to be present in trypanosomes, but were chosen as positive controls. Plasmid ID shows the identification number in our lab, when applicable. For Plasmid ID n.a. (not applicable) no plasmid was created due to failed PCR. PCR for cloning shows the concentration (in pmol) of DNA used for T4 DNA polymerase treatment, when performed during this project. When already created plasmids were used for protein over expression the person who created the original plasmid is named. Protein over expression indicates if the over expression was successful and Metabolomics assay refers to the appropriate chapter in this thesis.

3.2.3.1. S-adenosylmethionine synthetase, putative (G191)

The enzyme S-adenosylmethionine synthetase catalyses the formation of S-adenosylmethionine (AdoMet) from methionine and ATP, and is the starting enzyme in the methionine cycle (Figure 3.6).

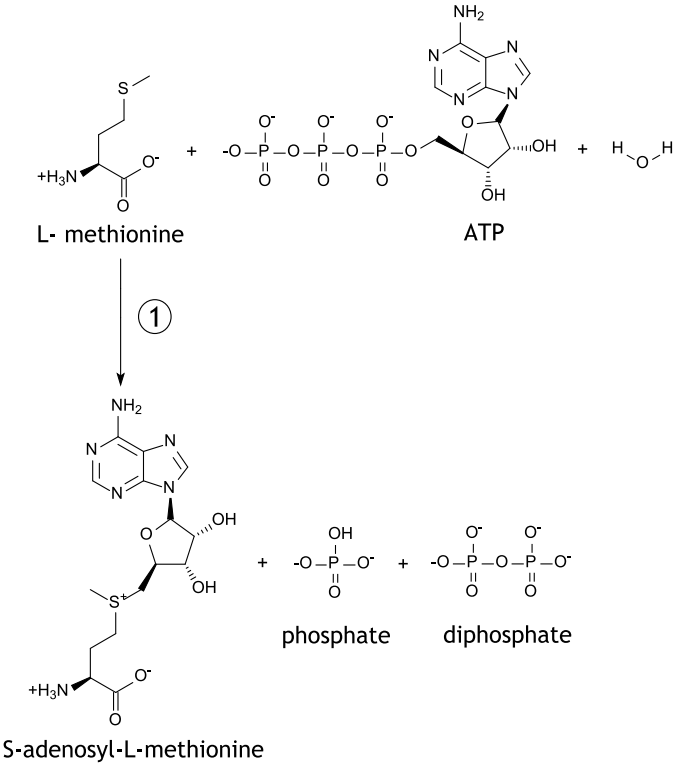


Figure 3.6: First step in the methionine cycle, adapted from Metacyc (http://metacyc.org). 1 S-adenosylmethionine synthetase (EC 2.5.1.6) catalyses the first step from L-methionine to S-adenosyl-L-methionine, using ATP as the adenosyl-group donor, leaving diphosphate and phosphate. Trypanosome genome shows nine genes coding for S-adenosylmethionine synthase.

S-adenosylmethionine is the key branch point in cell metabolism. Most of the intracellular methionine is converted into S-adenosylmethionine (Nozaki et al., 2005) and there are three known downstream reactions for AdoMet:

(1) AdoMet acts as the methyl group donor for most cellular methyltransferase reactions, in fact all biological methylation reactions with the exception of the methylation of homocysteine (Stipanuk, 2004).
(2) AdoMet can also be used for the formation of polyamines via decarboxylated AdoMet. It serves, in trypanosomes, as the aminopropyl group donor in the synthesis of polyamines, including Spermidine (Reguera et al., 2007).
(3) AdoMet can be recycled back to Methionine, via S-adenosyl homocysteine and homocysteine, a process known as the methionine cycle.

Activity of the reaction has been measured in cell extract of *T. brucei* and two isoforms of the enzyme are present in the bloodstream form (Bacchi and Yarlett, 1993). TritrypDB shows nine copies of a putatively annotated gene in *T. brucei* (Tb427.06.4840 / 4850 / 4860 / 4870/ 4880/ 4890/4900/ 4910/ 4920), but no experimental data with recombinant protein exists to confirm the correct annotation.

A gene encoding for S-adenosylmethionine synthetase was cloned from *T. b. brucei* strain 427 using the ligase independent cloning system pET30 Xa/LIC with primers specifically designed to create an overhang compatible with this system.
Primers GGTATTGAGGGTCGCATGTCCGTGCGCCAG (MB0729, fwd) and AGAGGAGAGTTAGAGCCCTACTGCACGTCACTAAGACC (MB0730, rev) were designed to produce the desired insert (introduced overhang is underlined).

The activity of recombinant S-adenosylmethionine synthetase was investigated using an *in vitro* assay combined with metabolite profiling. The recombinant protein was purified using an Immobilised Metal ion Affinity Chromatography (IMAC) protocol (Figure 3.7).

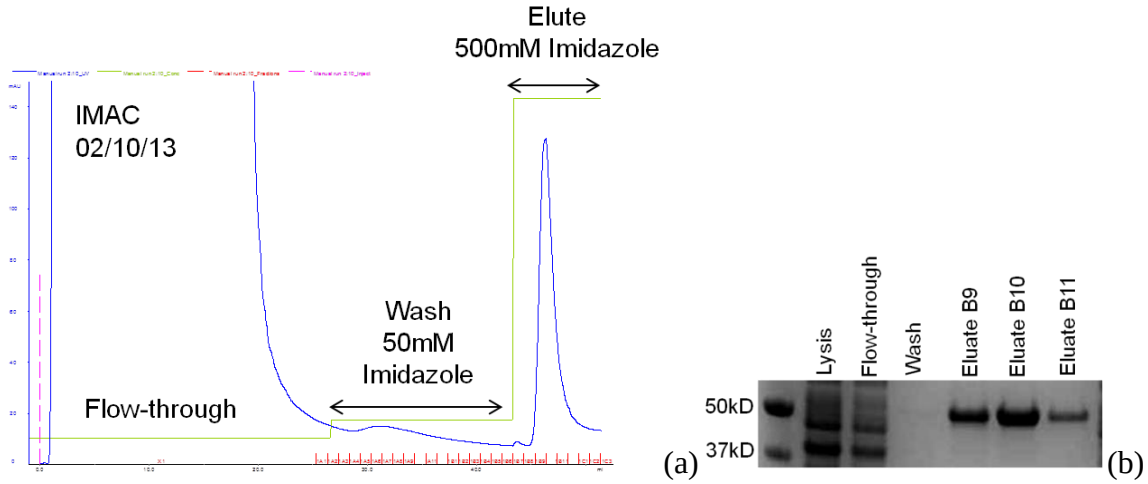


Figure 3.7: Purification of recombinant putative S-adenosylmethionine synthetase (SAM synthetase). SAM synthetase was heterologously expressed and purified using immobilised metal affinity chromatography (IMAC). (a) Purification profile using IMAC, with the protein of interest being eluted with 500mM imidazole. (b) Protein verification by SDS-PAGE**.**

The raw data was analysed using IDEOM (2.4.6). A heatmap was created in IDEOM using all identified basepeaks (Figure 3.8 (a)) and shows a minority of metabolites changing between the sample sets. Principal component analysis (PCA) of the dataset shows a less defined separation between control and treatment samples (Figure 3.8 (b)). The dataset showed a significant increase in S-adenosylmethionine in the sample set containing ATP (Figure 3.9 (a)). A low increase was also detected in the enzyme treated sample set

containing ADP (Figure 3.9 (a)). Whether that is because the commercially available ADP from Sigma was contaminated with ATP or because low levels of S-adenosylmethionine were present in the yeast extract is unknown.

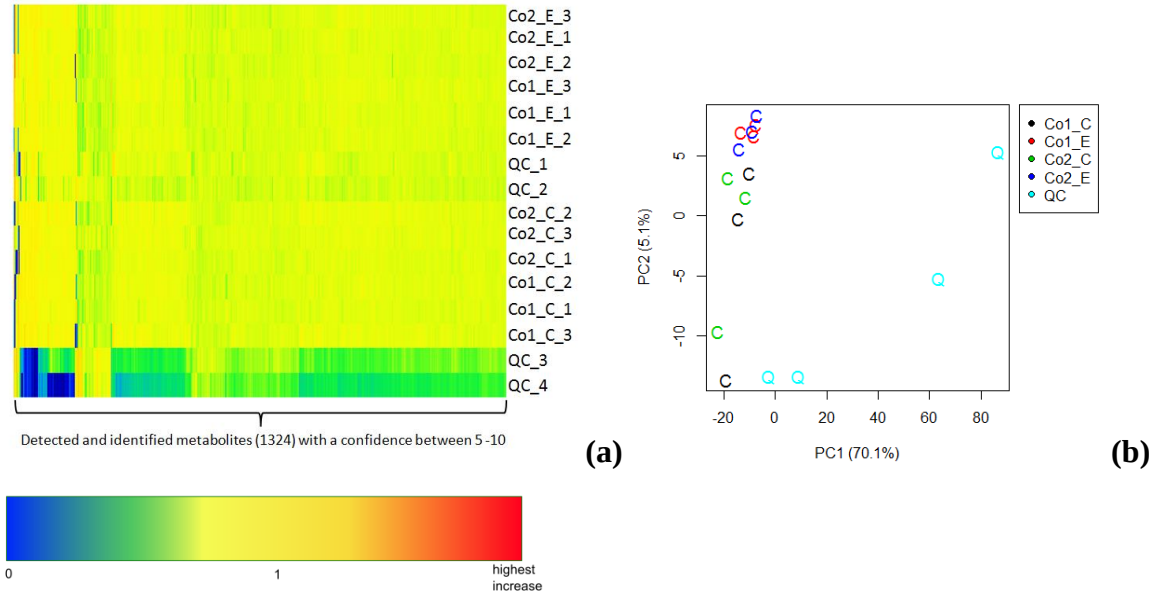


Figure 3.8: Heatmap and principal component analysis (PCA) of metabolomics dataset from SAM synthetase assay. (a) Heatmap created with the statistical software R in IDEOM from all identified metabolites with a confidence level between 5 and 10. The total number of detected metabolites was 1324. (b) PCA analysis of the same dataset. Abbreviations: Co1_C control sample of cofactor 1 metabolite mix, Co1_E enzyme treated sample of cofactor 1 metabolite mix. Co2_C / Co2_E cofactor 2 metabolite mix control (C) and enzyme treated (E) sample. In the heatmap the last digits representing the replicate number (1-3). Heatmap colours are represented in a gradient from dark blue (metabolite not present or significantly decreased, green (decreased, not significant), yellow (unchanged) and red (significantly increased).

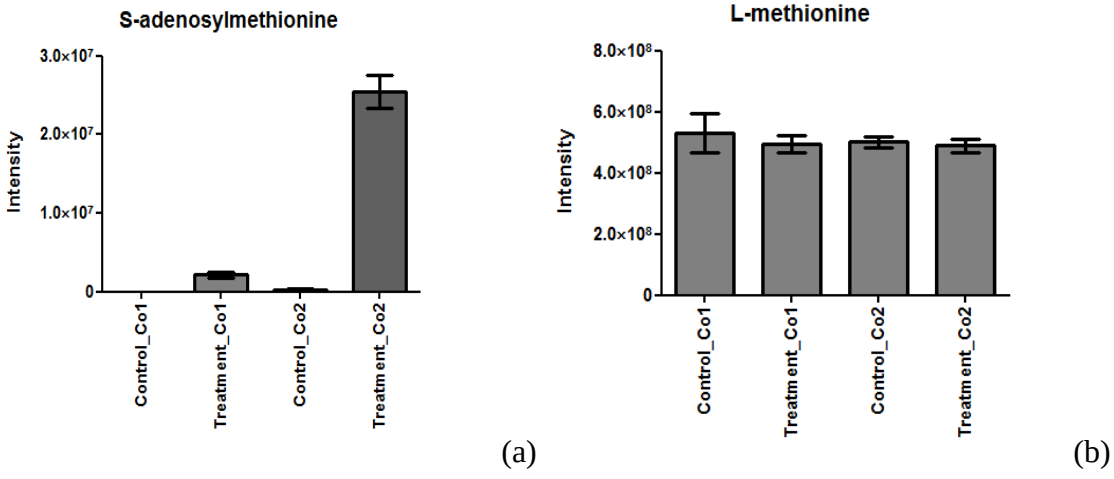


Figure 3.9: Peak intensity of S-adenosylmethionine and L-methionine. Assay was set up as described in 2.4.1. and analysed with IDEOM. (a) S-adenosylmethionine (the product of s-adenosylmethionine synthetase) and (b) L-methionine (the substrate of the reaction). Comparing the intensities of those two metabolites, it shows that L-methionine is highly abundant in the yeast extract which probably explains why the levels of L-methionine do not decrease in this experiment.

Unfortunately, no decrease in L-methionine levels was detected, however, as seen in Figure 3.9 (b), L-methionine was highly abundant in the yeast extract and it seems that the high levels of that compound masked the decreasing levels in the enzyme treated sample set containing ATP.

3.2.3.2. Citrate synthase, putative (G194)

Gene Tb427.10.13430 has been putatively identified as citrate synthase, which is the first enzyme in the citric acid cycle and synthesizes citrate from acetyl-CoA and oxaloacetate (Figure 3.10).

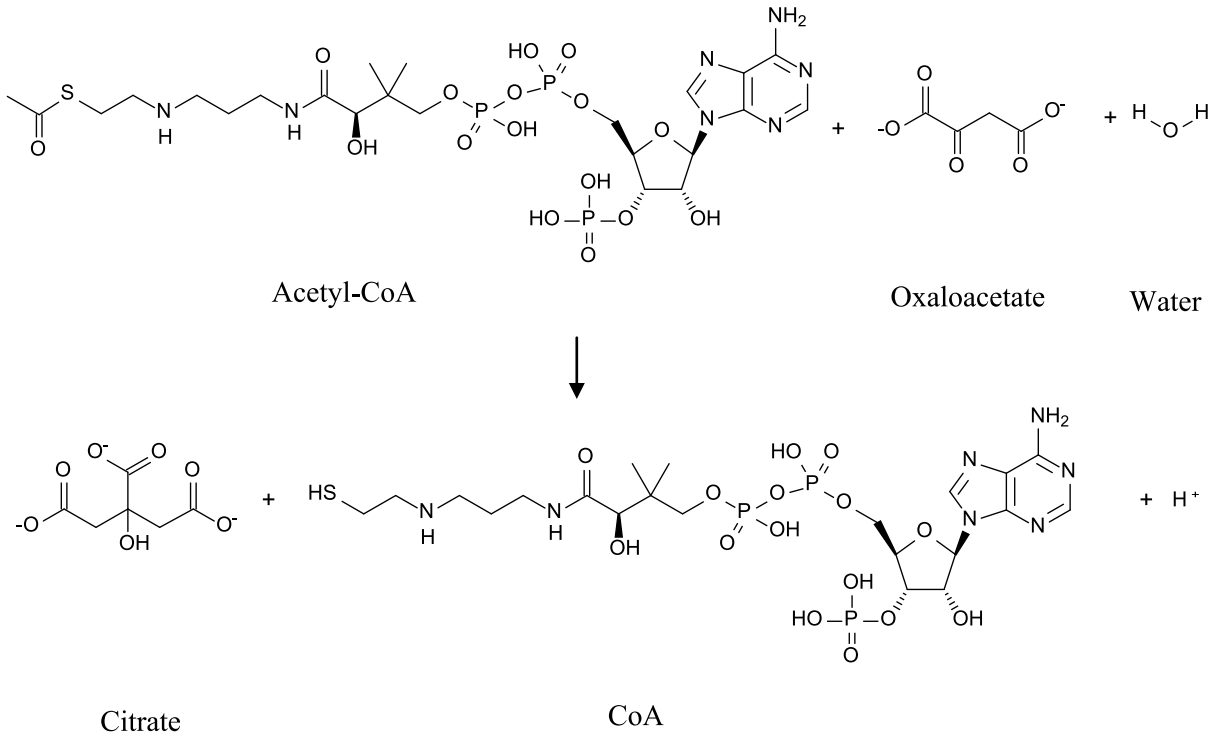


Figure 3.10: Reaction catalysed by Citrate synthase. Citrate and Coenzyme A are being produced from acetyl-CoA, oxaloacetate and water. Adapted from Metacyc.org.

Gene Tb427.10.13430 was cloned from *T. brucei* strain 427 using the ligase independent cloning system pET30 Xa/LIC with primers specifically designed to create an overhang compatible with this system.

Primers GGTATTGAGGGTCGCATGTGCATGCGTGCTCG (MB0739, fwd) and AGAGGAGAGTTAGAGCCCTACGCTATGTTGTACTTTGTG (MB0740, rev) were designed to produce the desired insert (introduced overhang is underlined). Protein was heterologously expressed in *E. coli* Rosetta (DE3) pLysS and purified using IMAC protocol described in method section (Chapter 2.3.2).

The *in vitro* assay combined with metabolite profiling showed very few changes in the metabolite mix between control and treatment samples. Heatmap and PCA analysis (Figure 3.11) show no pattern that would show sample separation from treatment *vs* control; however, a few changes indicated that there might be a reaction in the cofactor 2 treatment samples. The citrate synthase reaction, as shown in Figure 3.9., was not evident in this dataset. Due to the large sample variation (Figure 3.12 (b)), it is difficult to detect any possible changes in levels of citrate, but also the levels of acetyl CoA do not seem to change significantly between control and treatment samples (Figure 3.12 (a)).

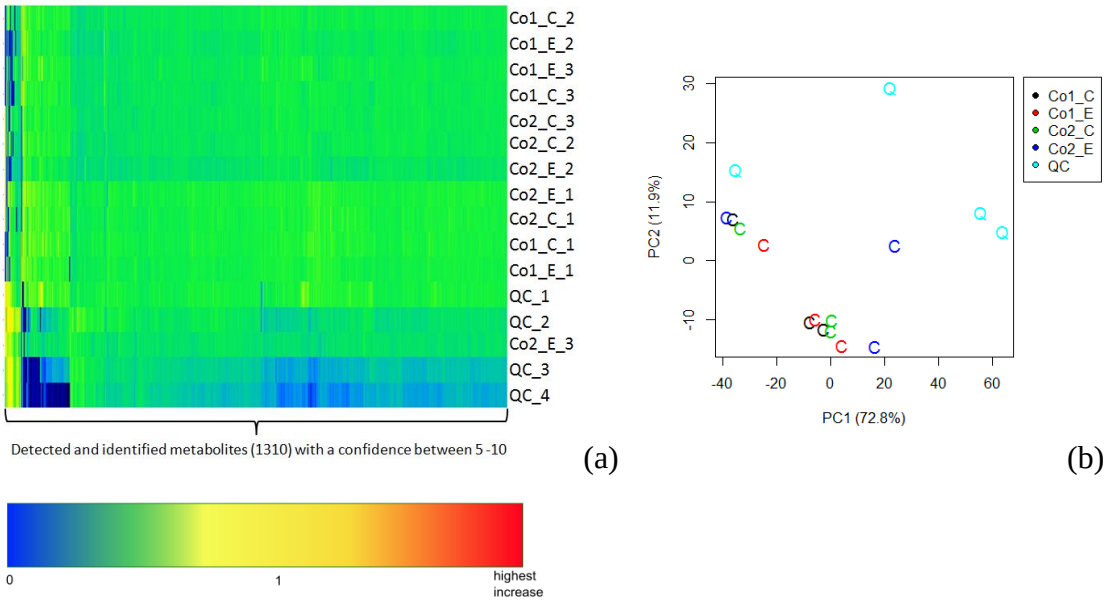


Figure 3.11: Heatmap and principal component analysis (PCA) of metabolomics dataset from citrate synthase assay. (a) Heatmap created with the statistical software R in IDEOM from all identified metabolites with a confidence level between 5 and 10. The total number of detected metabolites was 1310. Abbreviations: Co1 cofactor 1 metabolite mix, Co2 cofactor 2 metabolite mix with control (C) and enzyme treated (E) sample for each set and the last digits representing the replicate number (1-3). (b) PCA analysis created by the statistical software R using IDEOM, shows no clustering of sample sets (treatment *vs* control). Heatmap colours are represented in a gradient from dark blue (metabolite not present or significantly decreased, green (decreased, not significant), yellow (unchanged) and red (significantly increased).

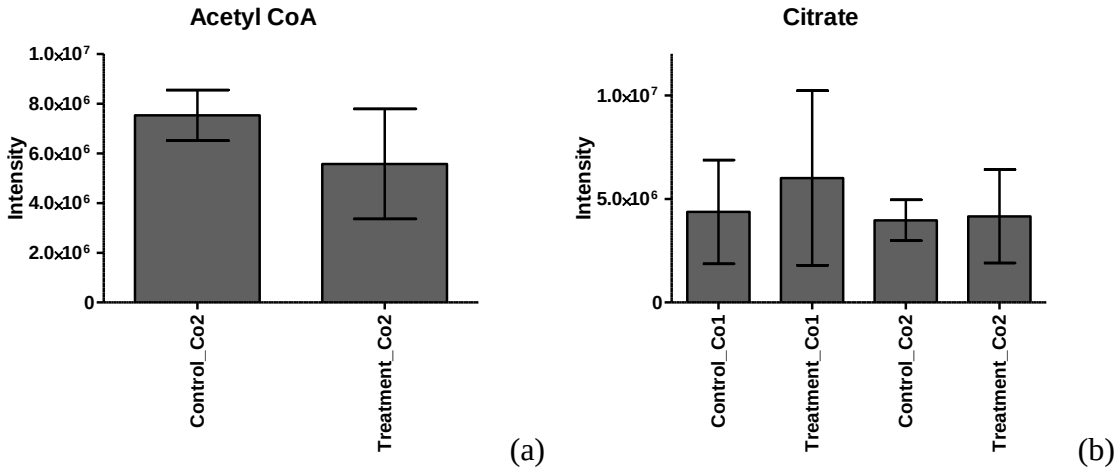


Figure 3.12: Peak intensities of metabolites predicted to change in the dataset according to putative annotation of enzyme. However, as seen in (a) levels of Acetyl-CoA (in samples containing cofactor 2 mix) do not seem to change significantly when sample mix is treated with enzyme compared to control sample and citrate levels are similar between control and treatment samples (b).

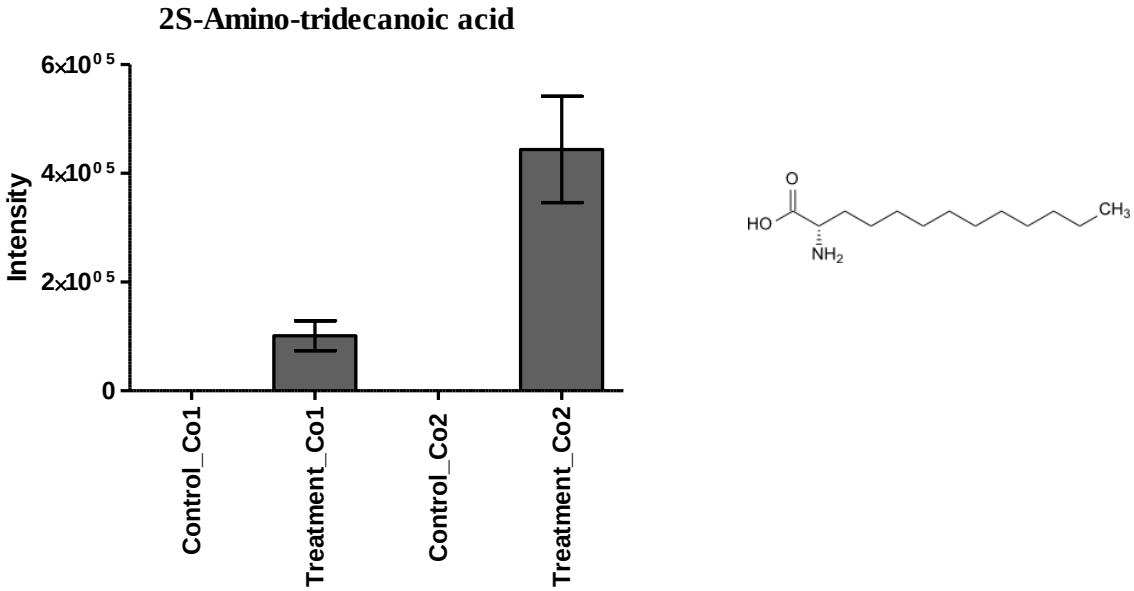


Figure 3.12 c: Peak intensity of significantly increased metabolite in G194 dataset putatively annotated as 2S-amino-tridecanoic acid. Changes in yeast extract with cofactor 1 mix, untreated (Control), treated (Treatment) with enzyme are represented on the left and changes in yeast extract with cofactor 2 mix, untreated, treated, are shown on the right. Structure of this compound is shown next to the changes.

The most significant change in this dataset was a compound with the mass of 229.20 and a retention time of 4.31 min, which has been putatively identified as 2S-Amino-tridecanoic acid (basepeak) with a confidence of 7 (Figure 3.12 (c)). However, 5 isomers are assigned to this compound (Table 3.5); with 2S-Amino-tridecanoic acid showing a retention time

(rt) closest to the predicted rt. This metabolite was only detected in the enzyme treated samples, with higher change detected in the cofactor 2 treated sample set (Figure 3.12 (c)).

Isomers	ppm error	predicted RT error
2S-Amino-tridecanoic acid	0.72	-7.70%
[FA amino(13:0)] 13-amino-tridecanoic acid	0.72	-28.10%
[FA amino(13:0)] 2R-aminotridecanoic acid	0.72	-7.90%
[FA amino(13:0)] 2-amino-tridecanoic acid	0.72	-7.90%
Capryloylcholine	0.72	-49.40%

Table 3.5: Isomers of detected compound showing significant increase in treatment samples. Putative identification was 2S-Amino-tridecanoic acid, which showed the lowest retention time error to the predicted retention time.

2S-Amino-tridaconic acid is an amino fatty acid belonging to the 13C-carbon saturated fatty acids. Searches in metacyc.org, KEGG and PubMed have not indicated the possible existence of a pathway containing this metabolite in trypanosomes. Therefore, it is possible to give a possible function for this protein on this one significant increase found in the dataset.

3.2.3.3. Deoxyhypusine synthase, putative (G158)

Tb927.10.2750 is putatively identified as a deoxyhypusine synthase, which catalyses the spermidine-dependent modification of hypusine in a lysine residue for the essential translation factor elF5A. This reaction is performed in two steps, the first step deoxyhypusine synthase transfers the butylamine moiety from spermidine to a specific lysine residue of the of the eIF5A precursor protein with NAD+ as a cofactor (Joe et al., 1995). However, if the elF5A precursor protein is absent, the reaction releases 1-pyrroline. Decreasing levels in Spermidine with corresponding increasing levels in 1-pyrroline could, in theory, be detected using the LC-MS platform used in this study. Gene Tb927.10.2750 was cloned from *T. brucei* strain 427 using the ligase independent cloning system pET30 Xa/LIC with primers specifically designed to create an overhang compatible with this system.

Primers GGTATTGAGGGTCGCATGGCTGAGTTGGCCAAGAG (MB0639, fwd) and AGAGGAGAGTTAGAGCCTCACGAGCGGATATTCTCCT (MB0640, rev) were designed to produce the desired insert (introduced overhang is underlined) and cloning was successful and confirmed by sequencing.

As described above the changes I would have expected to see during this experiment, if this protein is indeed a deoxyhypusine synthase, are decreasing levels of Spermidine and NAD+ and increasing levels of 1-pyrroline. However, only a minimal change in spermidine levels was observed during the *in vitro* metabolite profiling (Figure 3.12. (a)). The sample set containing yeast metabolites and Cofactor mix 1 (which includes NAD+), treated with enzyme showed a slight decrease in levels of spermidine. However, 1-pyrroline and NAD+ were not detected. These results are not conclusive enough to identify the gene Tb927.10.2750 as an deoxyhypusine synthase coding gene.

After this experiment was performed it was shown that this protein is indeed a deoxyhypusine synthase but also that this enzyme's activity is increased 3000-fold by forming a heterotetramer with a catalytically dead paralog (Nguyen et al., 2013). This would explain why the results were inconclusive using the metabolite profiling approach.

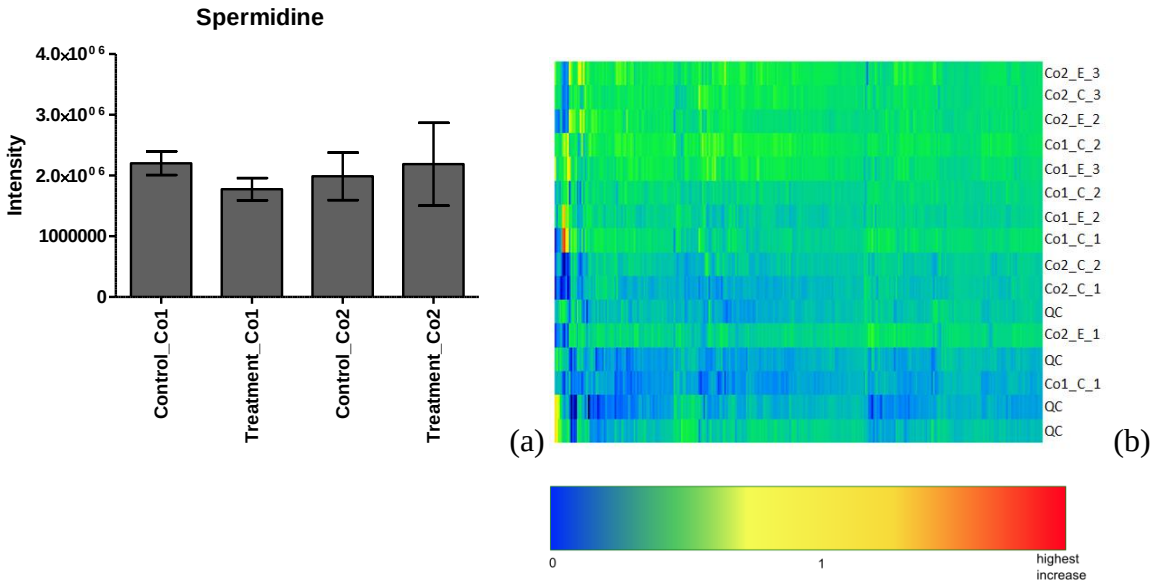


Figure 3.13: Spermidine peak intensities and heatmap of metabolomics dataset from deoxyhypusine synthase assay. (a) Average peak height of spermidine from three sample replicates with standard deviation and (b) Heatmap of untargeted metabolite profiling approach with yeast extract and two cofactor mixes (3 replicates) of Deoxhypusine synthase. No significant changes in metabolite levels between control and treatment samples were detected in this dataset. Abbreviations are: Co1 for cofactor 1 / metabolite mix, Co2 for cofactor 2 / metabolite mix with control (C) and enzyme treated (E) sample for each set and the last digits representing the replicate number (1-3). Heatmap colours are represented in a gradient from dark blue (metabolite not present or significantly decreased, green (decreased, not significant), yellow (unchanged) and red (significantly increased).

As shown in the heatmap of this experiment (Figure 3.13) there seem to be some changes in the data set. However, after analysis it became clear that those changes demonstrate the variance between samples more than sample groups (control/treatment). This was evident from replicates not showing similar intensities. As IDEOM uses the average peak intensity

to compare control to treatment, a metabolite not detected in one replicate could indicate a change when indeed it is just sample variation.

3.2.3.4. Deoxyribose-phosphate aldolase, putative (G159)

The gene Tb927.7.5680 codes for a protein that is putatively annotated as deoxyribose-phosphate aldolase, an enzyme which catalyses the following reaction:

2-deoxy-D-ribose 5-phosphate → D-glyceraldehyde 3-phosphate + acetaldehyde

To investigate the function of this protein, the gene Tb927.7.5680 was cloned from *T. b. brucei* strain 427 using target gene specific primers and a ligase independent cloning system (as described in the methods section), using primers GGTATTGAGGGTCGCATGACCGACCTTCACATGAG (MPB0641, fwd) and AGAGGAGAGTTAGAGCCTTAGTATTTACTGCGGGAGC (MPB0642, rev), to create vector specific overhang (Introduced overhang is underlined). The untargeted metabolomics approach showed no significant changes regarding any potential enzyme activity. The substrate and product have not been detected, although previous data sets have shown those metabolites to be present in the yeast extract. Although the PCA analysis (Figure 3.14 (b)) shows a grouping of samples treatment *vs* control, the dataset showed no significant changes in individual metabolite levels between those groups (Figure 3.14 (a)).

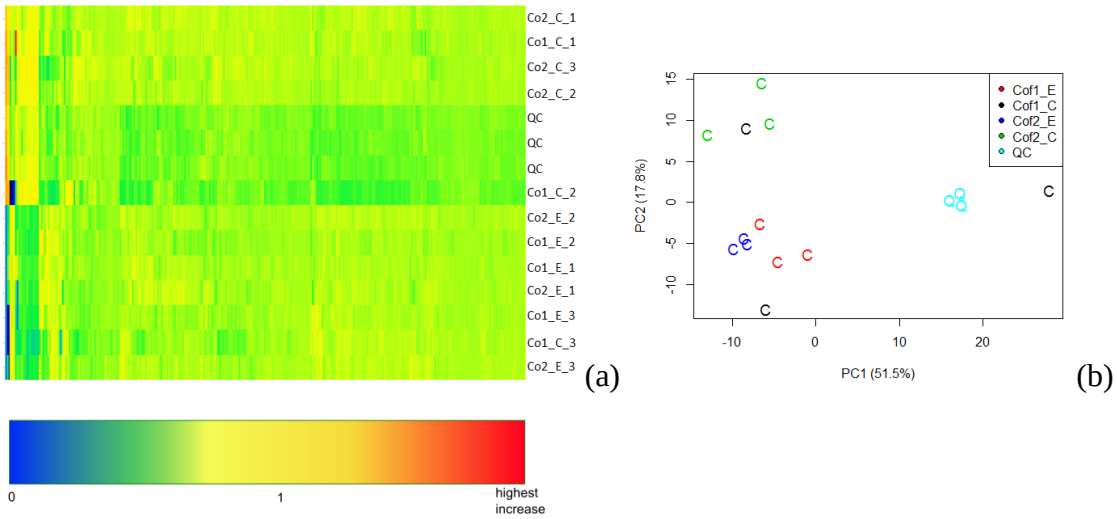
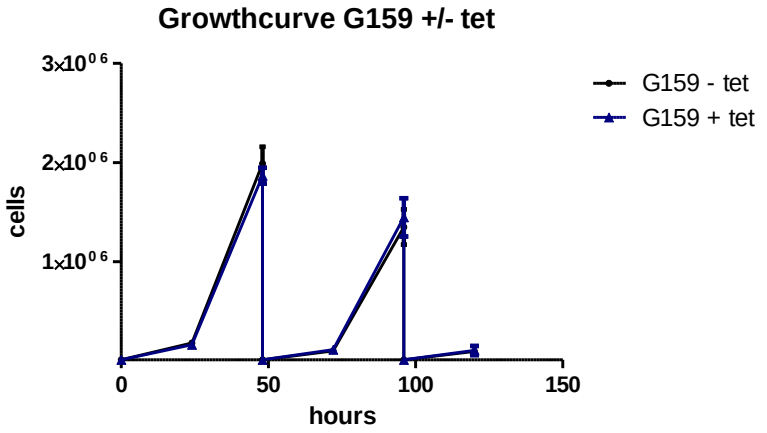


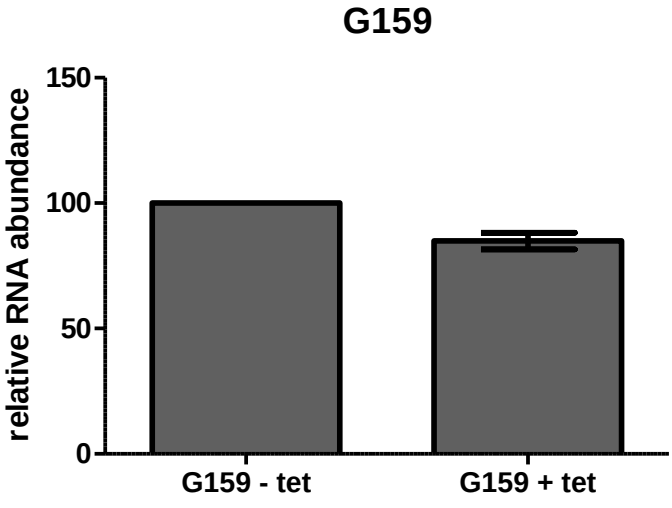
Figure 3.14: Heatmap and principal component analysis (PCA) of metabolomics dataset from deoxyribose-phosphate aldolase assay. (a) Heatmap of untargeted metabolite profiling approach with yeast extract and two cofactor mixes (3 replicates) of deoxyribose-phosphate aldolase. (b) PCA analysis of same data set shows clustering of treatment *vs* control samples (treatment samples in red and dark blue). Abbreviations are: Co1 for cofactor 1 / metabolite mix, Co2 for cofactor 2 / metabolite mix with control (C) and enzyme treated (E) sample for each set and the last digits representing the replicate number (1-3). Heatmap colours are represented in a gradient from dark blue (metabolite not present or significantly decreased, green (decreased, not significant), yellow (unchanged) and red (significantly increased).

To establish if this protein is essential for bloodstream form trypanosomes, a deoxyribose-phosphate aldolase^{RNAi} cell line (G159^{RNAi}) was created and the transcription of the gene knocked down by RNA interference.

Induction of knock down with 1 µg ml^{-1} and 5 µg ml^{-1} tetracycline showed no change in growth phenotype (see Figure 3.15 (a)). Reverse transcription with RT-PCR showed only a knockdown effect of about 10% (Figure 3.15 (b)). Therefore it could not be determined whether this protein is essential for cell survival.



(a)



(b)

Figure 3.15: Growth curve of G159^{RNAi} cell line (a) and relative RNA abundance in G159^{RNAi} cell line, as determined by rt RT-PCR (b)
(a) Culture of G159^{RNAi} , induced every 24 h with 1 µg ml^{-1} tet (+) and without tet (-) , was monitored for six days for differences in growth rate. As seen, no differences in growth could be detected. Cell flasks were set up in triplicate.
(b) Two cultures of G159^{RNAi} were set up, one was induced with 5 µg ml^{-1} tetracycline (+tet) every 24 hours. Samples were taken (5 x 10^{7} cells) at timepoint 48h, 72h and 96h and RNA extracted and relative RNA abundance of transcript of interest was assessed by rt RT-PCR (in triplicate). Timepoints were combined for graph as there were no changes in RNA abundance between different timepoints. Protein GPI-8 was used as standard.

3.2.3.5. NAD^+ synthase, putative (G162)

NAD^+ synthase catalyses the last step in the NAD^+ biosynthesis pathway and the reaction involves either the transfer of an amino group from glutamine or ammonia to form NAD^+ from nicotinic acid adenine dinucleotide (NaAD or Deamido-NAD^+) (Ozment et al., 1999). The reactions are shown in Figure 3.16 (glutamine dependent NAD^+ synthase) and Figure 3.17 (ammonium dependent NAD^+ synthase).

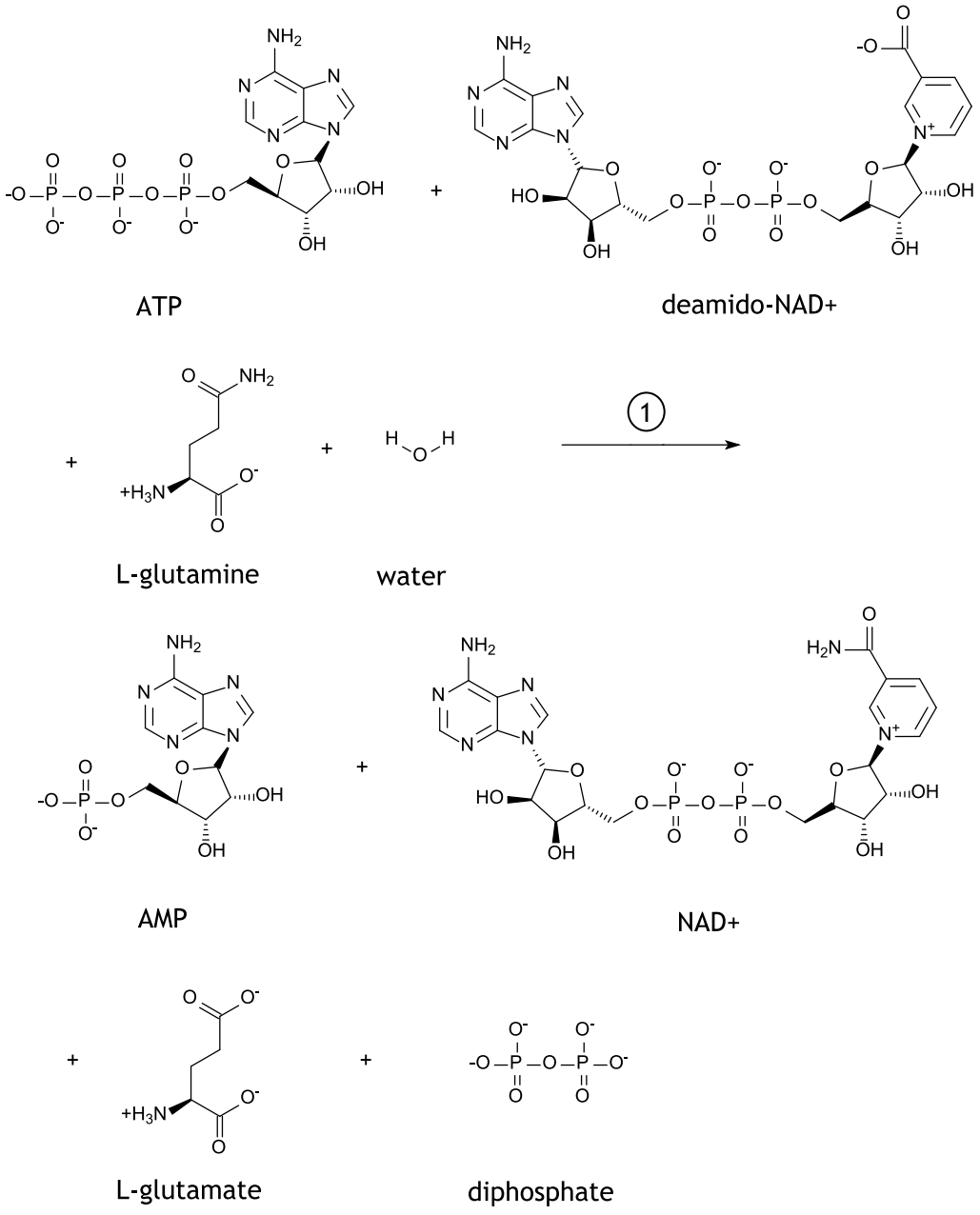


Figure 3.16: Reaction of L-glutamine dependent NAD^+ synthase, adapted from metacyc.org. In this reaction an amino-group is transferred from L-glutamine to deamido-NAD+ to form L-glutamate and NAD+.

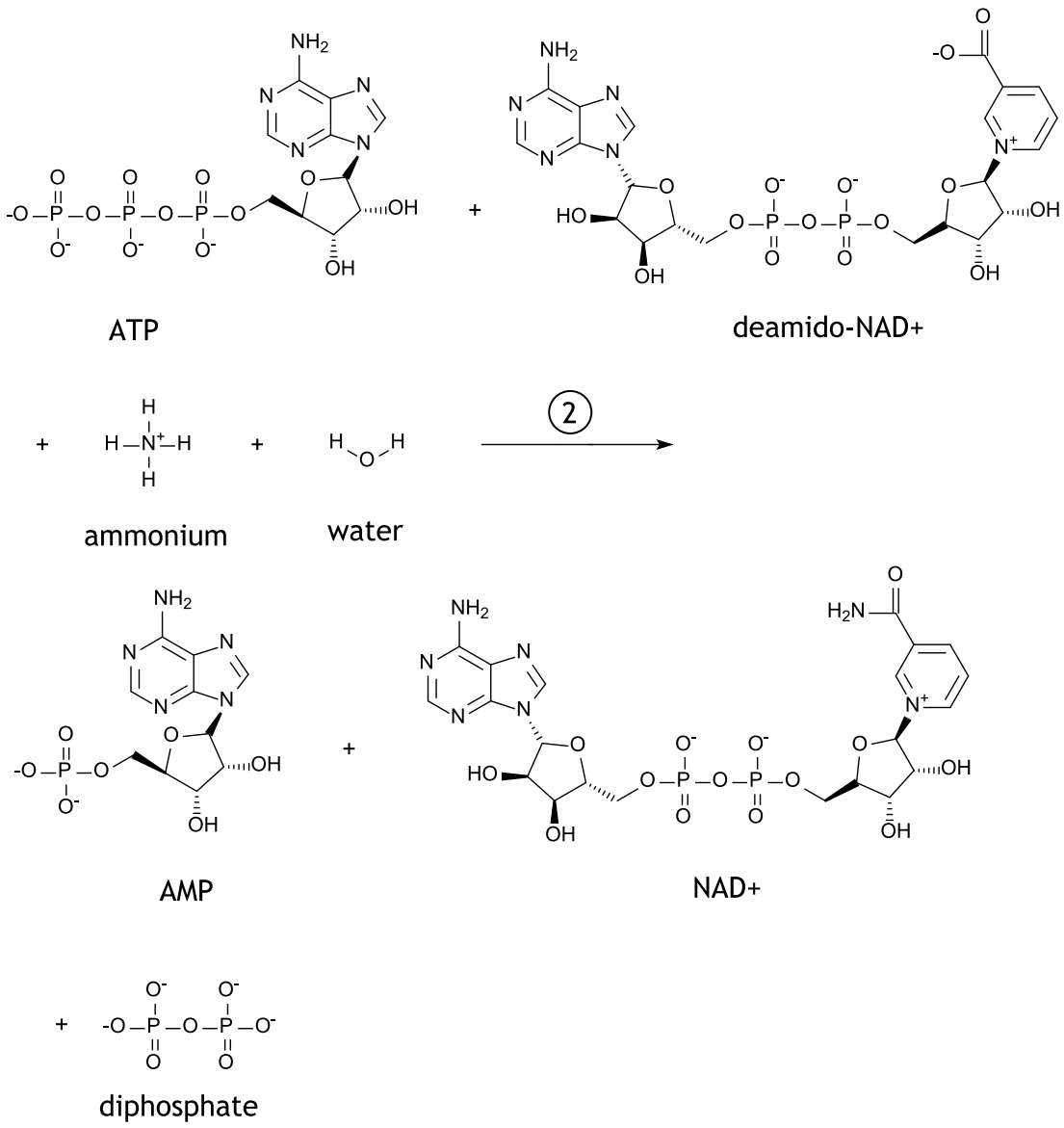


Figure 3.17.: Reaction of ammonium dependent NAD$^+$ synthase, adapted from metacyc.org. In this reaction ammonium and deamido-NAD$^+$ form NAD$^+$.

E. coli favours the ammonia dependent reaction (Spencer and Preis, 1966), while eukaryotic NAD$^+$ synthase seems to be glutamine dependant (Wojcik et al., 2006). In trypanosomes, gene number Tb11.01.6500 is putatively annotated as an NAD$^+$ synthase and was investigated for its biological function using in vitro metabolite profiling and targeted enzyme assays. To study the function of the encoded protein from gene Tb11.01.6500, it was cloned from *T. brucei* strain 427 using target gene specific primers and a ligase independent cloning system (as described in the methods section), using primers GGTATTGAGGGTCGCATGCCGAAGGAGCCCATTCT (MPB0647, fwd) and AGAGGAGAGTTAGAGCCTACAGGTTCACAATACCGT (MPB0648, rev), to create vector specific overhang (Introduced overhang is underlined). Over-expression was

achieved using 1 litre cultures of transformed *E. coli* Rosetta pLysS cells as well as *E. coli* BL21 (DE3). The over-expression of the putatative NAD^+ synthase showed consistently high amounts of protein produced, but purification, using IMAC, showed contamination in the eluate (see Figure 3.18).

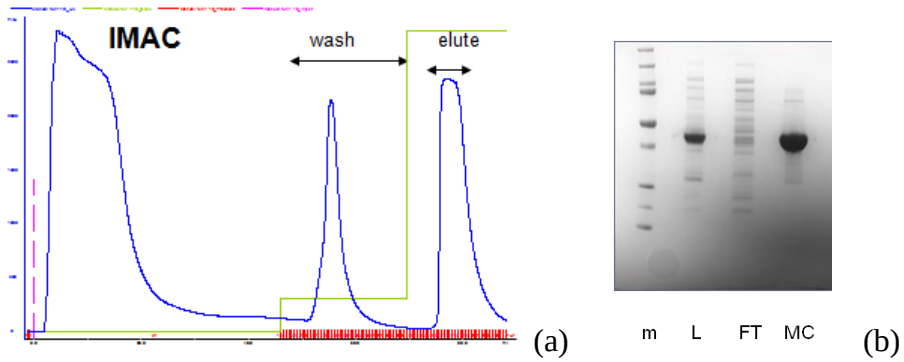


Figure 3.18: Purification of recombinant putative NAD^+ synthase. NAD^+ synthase was heterologously expressed and purified using immobilised metal affinity chromatography (IMAC). (a) Purification profile of recombinant protein G162 over-expressed in *E. coli* Rosetta pLysS, using IMAC, washed with 50 mM imidazole and eluted with 500 mM imidazole and (b) protein verification by SDS-PAGE, m=Marker, L=Lysis, FT= Flowthrough and MC=pool of eluate.

Heatmap and PCA are shown in Figure 3.19 and show little variance between the samples, except for the QC samples. However, the pattern of decreasing metabolites throughout the run indicates that the volume of QC sample was too low and sample could not be picked up.

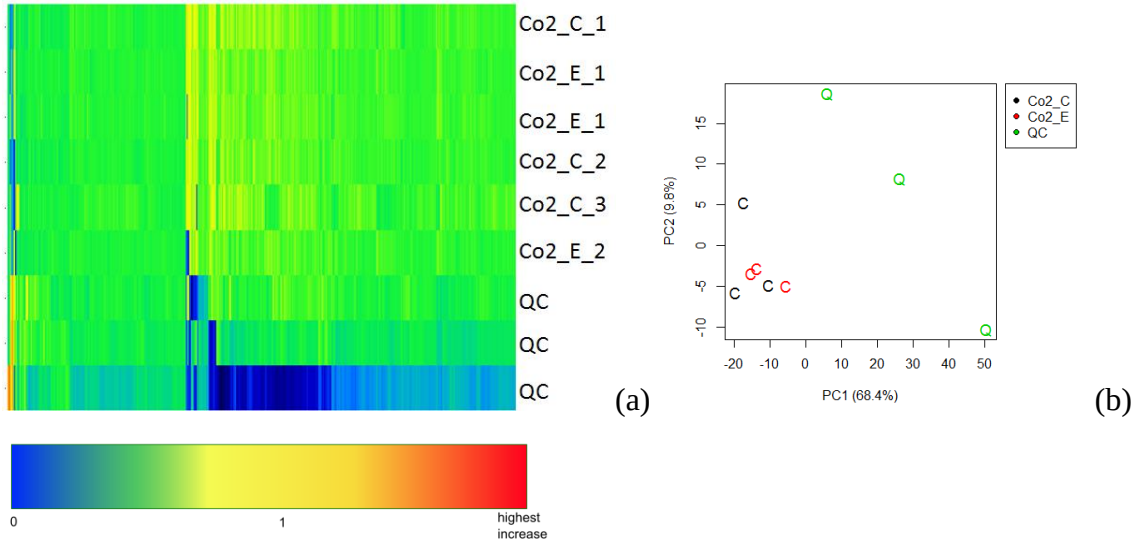


Figure 3.19: Heatmap and principal component analysis (PCA) of metabolomics dataset from NAD^+ synthase assay. (a) Heatmap of untargeted metabolite profiling approach with yeast extract and cofactor mix 2 (3 replicates). The visible changes occur in the quality control (QC) samples, which are run at the beginning, in the middle and at the end of the run, with decreasing detected metabolites from QC1 to QC 3. Those changes are most likely due to sample degradation. (b) Correlation between sample groups (control/treatment) was analysed by PCA. No clear separation is apparent between the two groups. Heatmap colours are represented in a gradient from dark blue (metabolite not present or significantly decreased, green (decreased, not significant), yellow (unchanged) and red (significantly increased).

Metabolic profiling by mass spectrometry did show decreasing levels of deamido-NAD^+ or L-glutamine, however those changes do not seem to be significant (Figure 3.20). ATP and NAD^+ were not detected in the IDEOM spreadsheet.

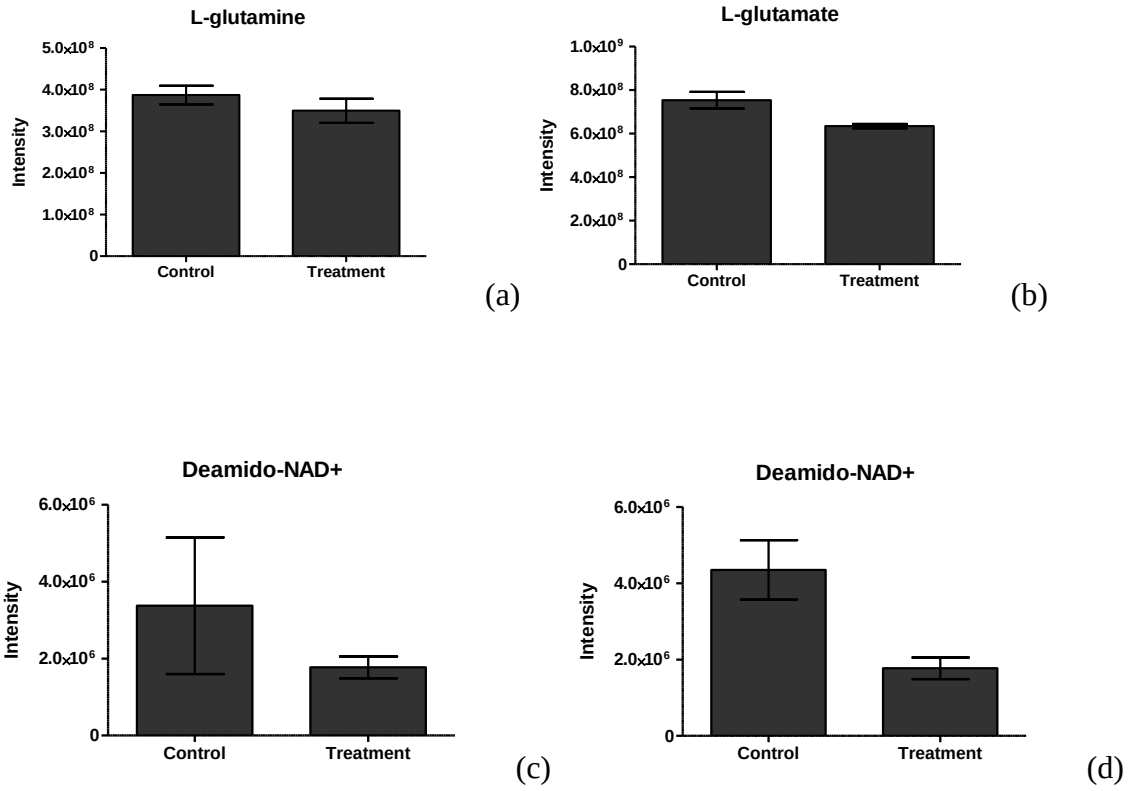


Figure 3.20: Average peak intensities of L-glutamine, L-glutamate and Deamido-NAD^+ in metabolomics dataset. Detected average intensities of metabolites, assumed to be involved in reaction, of cofactor 2 treatment (Co2_E) and control (Co2_C) samples (three replicates). Although it appears that L-glutamine levels go slightly down in treatment samples (a), those changes are not significant. L-glutamate levels show no corresponding increase (b). Deamido-NAD^+ (c) also shows no significant decrease in intensity, however, when one control sample, showing significantly lower peak intensity than the other samples, was removed, decrease of Deamido-NAD^+ seems to be significant (d). 3 replicates were used in this study.

As the detected changes in the metabolomics dataset would not immediately point to this enzyme being a NAD^+ synthase, a spectrophotometric enzyme assay was performed to determine the activity of the purified protein using an adapted coupled enzyme assay from Wojcik *et al* (2006). Instead of alcohol dehydrogenase (suggested enzyme for coupled reaction), glucose dehydrogenase was used. Protein purified from Rosetta pLysS seemed to produce NAD^+, while protein produced in BL21 did not.

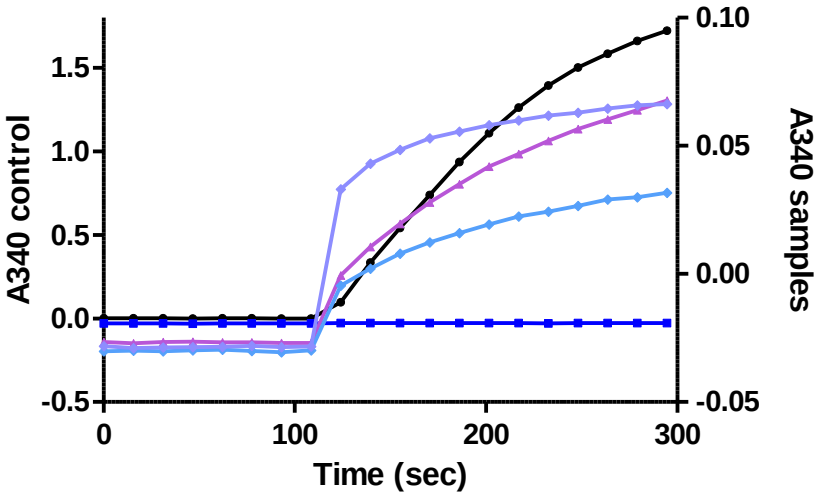


Figure 3.21: **Results of NAD$^+$ -synthase assay on spectrophotometer.** A glutamine-dependent enzyme assay was performed, as described by Wojcip *et al* (2006). Positive control with 20mM NAD$^+$ was plotted against left y-axis and is marked in black. Negative control, enzyme assay reaction mix 1 (as described in method section) without enzyme is shown in dark blue. Coupled enzyme assay, reaction mix 1 and glucose dehydrogenase assay is shown in triplicate (light blue, light purple and pink) and plotted against the right y-axis. Glucose dehydrogenase was added to all samples at 120 seconds.

As shown in Figure 3.21, the increase of absorbance was minimal in the NAD$^+$ synthase treated samples. However, although minimal changes in absorbance indicate that gene Tb11.01.6500 is indeed an NAD$^+$ synthase, it is not clear if the enzyme is an ammonium dependent or glutamine hydrolysing NAD$^+$ synthase. A spectrophotometric enzyme assay for ammonium dependent NAD$^+$ synthase showed no activity when performed with protein purified from *E. coli* BL 21 (DE3). It is likely though, that the trypanosome NAD$^+$ synthase is ammonium dependent, but that L-glutamine can still act as an amino group donor, albeit not as efficiently (Spencer and Preiss, 1967). If so, it would explain the minimal changes in the glutamine dependent enzyme assay. A bioinformatics search on the predicted protein sequence (EFICAz$^{2.5}$) suggested that the gene Tb11.01.6500 codes for the ammonium dependent enzyme. A further Pfam (http://pfam.xfam.org/) search for active domains revealed that the gene Tb11.01.6500 only contains one domain associated for NAD$^+$ synthase activity. This result is consistent with *E. coli* NAD$^+$ synthase. Human NAD$^+$ synthase on the other hand shows two domains. The 'CN hydrolase' domain appears to be necessary for the glutamine dependant activity. This domain is missing from trypanosomes NAD$^+$ synthase, however, it is found in trypanosome gene Tb927.9.1960, which is putatively annotated as a nitrilase.

3.2.3.6. Aldo/keto reductase, putative (G196)

The aldo/keto reductase superfamily includes NADPH dependant oxidoreductases (Bohren *et al.,* 1989). Gene Tb427tmp.02.3040 was cloned from *T. b. brucei* strain 427 using the ligase independent cloning system pET30 Xa/LIC with primers specifically designed to create an overhang compatible with this system.

Primers GGTATTGAGGGTCGCATGGACCGTATTCCATATTTGG (MB0745, fwd) and AGAGGAGAGTTAGAGCCTTAATCTATCGTGTTGCTATGCC (MB0746, rev) were designed to produce the desired insert (introduced overhang is underlined).

Protein was heterologously expressed in *E. coli* Rosetta (DE3) pLysS and purified using IMAC protocol described in method section (2.3.2). The recombinant *T. brucei* enzyme did not show any significant changes in the dataset that could lead to function of this enzyme. Heatmap and PCA analysis (Figure 3.22) showed no difference in control *vs* treatment samples. NADPH dependent changes should have been seen in cofactor mix 2, however, NADPH or NADP were not detected.

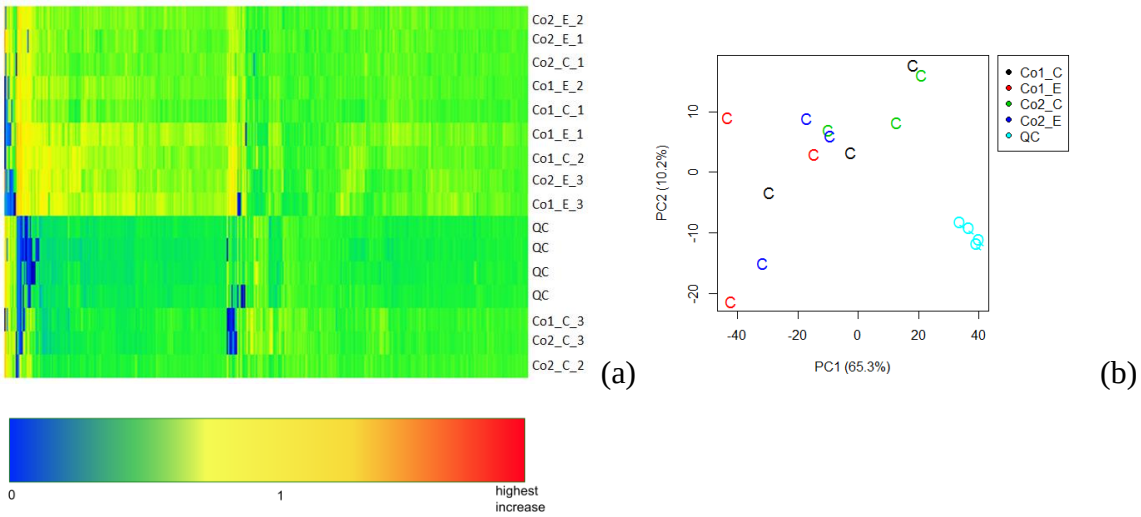


Figure 3.22: Heatmap and principal component analysis (PCA) of metabolomics dataset from aldo/keto reductase assay. (a) Heatmap created from dataset using R. Changes in metabolite levels show the variance between samples as levels vary within the replicates more than between the sample groups. (b) PCA analysis shows separation between the individual samples, indicating random changes between the samples. Abbreviations are: Co1 for cofactor 1 / metabolite mix, Co2 for cofactor 2 / metabolite mix with control (C) and enzyme treated (E) sample for each set and the last digits representing the replicate number (1-3). Heatmap colours are represented in a gradient from dark blue (metabolite not present or significantly decreased, green (decreased, not significant), yellow (unchanged) and red (significantly increased).

To establish if this protein is essential for bloodstream form trypanosomes an Aldo/keto reductase^{RNAi} cell line (G196^{RNAi}) was created and the transcription of the gene knocked down by RNA interference. Induction of knock down with 1 µg ml^{-1} or 5 µg ml^{-1} tetracycline showed no change in growth phenotype. Reverse transcription with RT-PCR showed a knockdown effect of about 40% (Figure 3.23).

No further studies were performed, as it was concluded that 40% knock down is not enough to establish if the protein is essential and therefore further metabolomics studies seemed uninformative.

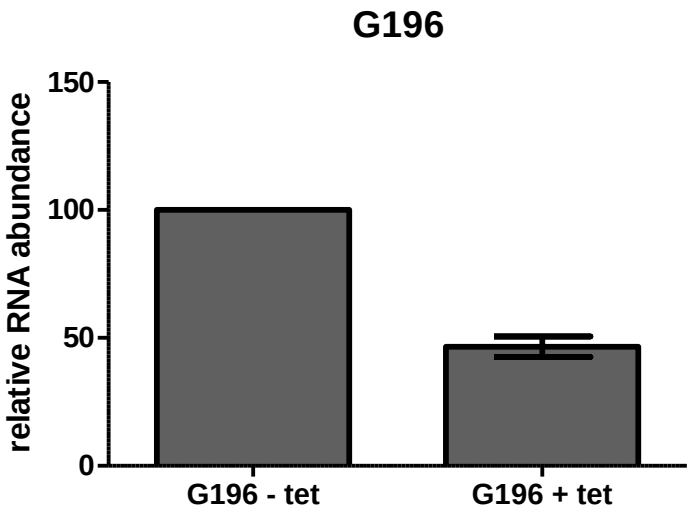


Figure 3.23: Relative RNA abundance in G196^{RNAi} cell line, as determined by rt RT-PCR
Two cultures of G196^{RNAi} were set up, one was induced with 5 µg ml^{-1} tetracycline (+tet) every 24 hours. Samples were taken (5 x 10^7 cells) at timepoint 48h, 72h and 96h, RNA extracted and relative RNA abundance of transcript of interest was assessed by rt (reverse transcription)RT-PCR (in triplicate). Timepoints were combined for graph as there was no change in RNA abundance between different timepoints. Protein GPI-8 was used as standard.

3.2.3.7. Aspartate carbamoyltransferase, putative (G197)

Aspartate carbamoyltransferase catalyses the first step in the pyrimidine biosynthetic pathway. It belongs to the enzyme class of transferases and the reaction fuses L-aspartate with a carbamoyl-group to form N-carbamoyl-L-aspartate (Figure 3.24). In bloodstream form trypanosomes, pyrimidine biosynthesis from glucose has been observed when 50 % U-^{13}C-labelled glucose was added to the growth medium (Creek et al., 2015).

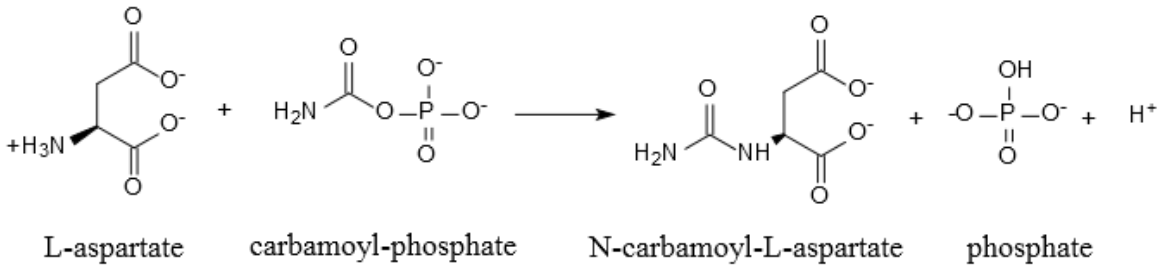


Figure 3.24: Reaction of aspartate carbamoyltransferase. Adapted from metacyc.org

Gene Tb427.05.3820 was cloned from *T. brucei* strain 427 using the ligase independent cloning system pET30 Xa/LIC with primers specifically designed to create an overhang compatible with this system.

Primers GGTATTGAGGGTCGCATGGCGGAGCTGCAACCTG (MB0749, fwd) and AGAGGAGAGTTAGAGCCTTAGGCGAGAACACTATAAAG (MB0750, rev) were designed to produce the desired insert (introduced overhang is underlined).

Protein over-expression was achieved consistently with high yield in both *E. coli* Bl21 (DE3) and Rosetta (DE3) pLysS. However, Rosetta (DE3) pLysS cells were used for this study and 1 litre culture was induced with 1mM IPTG overnight at 18°C and protein purified using IMAC (Figure 3.25).

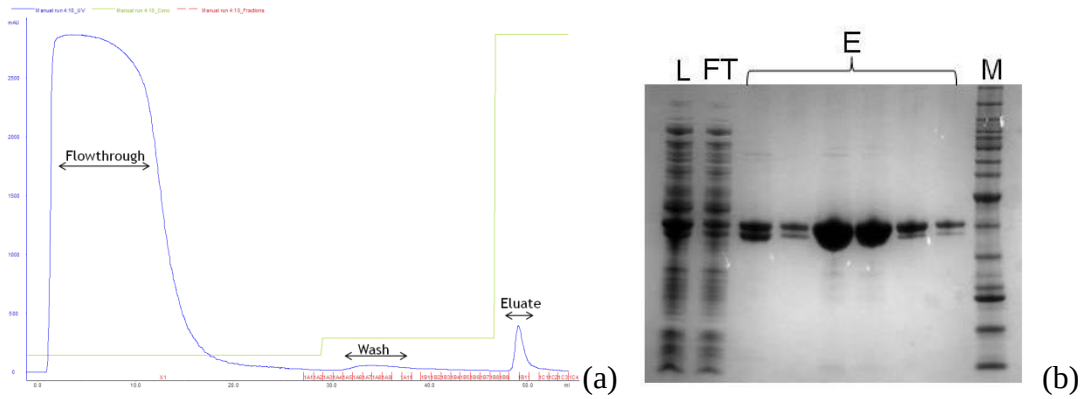


Figure 3.25: Purification of recombinant putative aspartate carbamoyltransferase. Aspartate carbamoyltransferase was heterologously expressed and purified using immobilised metal affinity chromatography (IMAC). (a) Purification profile of recombinant protein using IMAC, protein was washed with 50 mM imidazole and eluted with 500 mM imidazole. (b) Protein verification by SDS-PAGE, m=Marker, L=Lysis, FT= Flowthrough and E= eluates.

In vitro assay combined with metabolite profiling showed no differences between the control and treatment samples as shown in the heatmap and PCA (Figure 3.26). One Control sample (Co1_C_1) showed irregularity to the rest (see Figure 3.26 (a)). However, when this sample was excluded from analysis, no significant changes were seen.

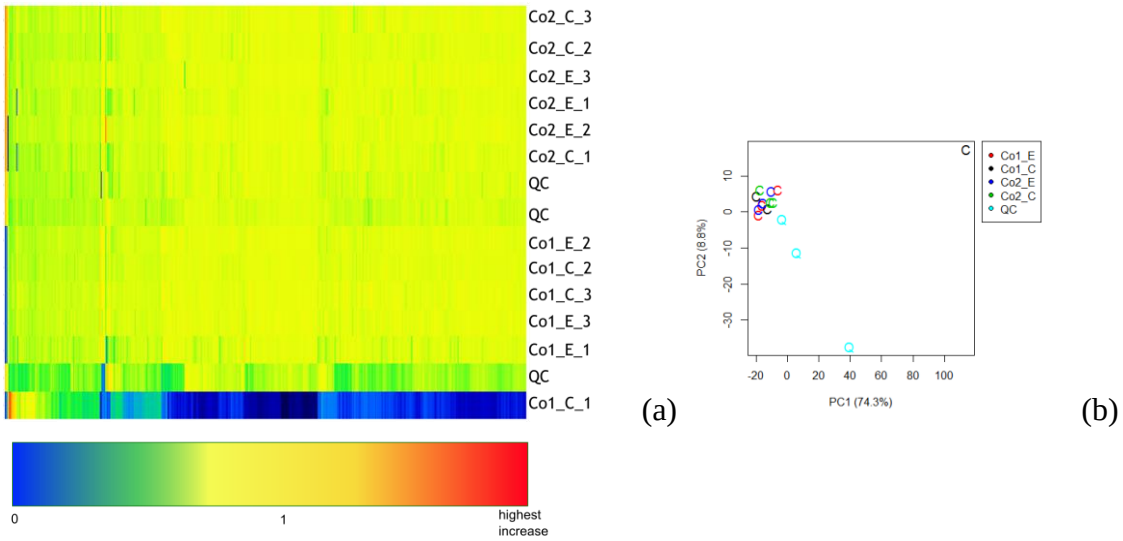
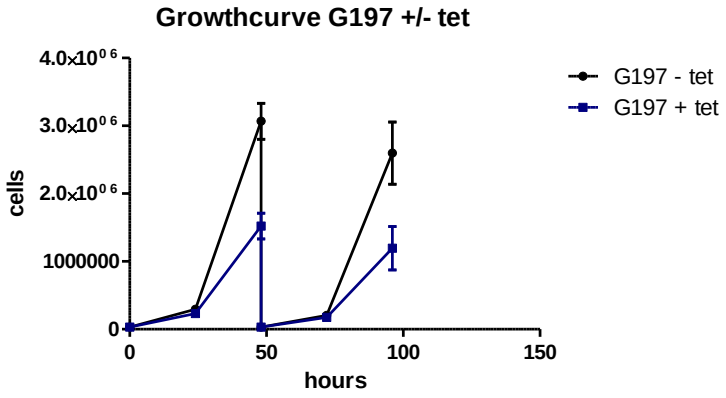
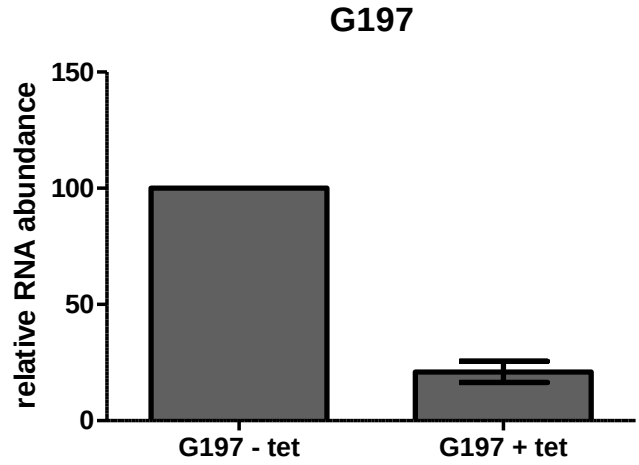


Figure 3.26: Heatmap and principal component analysis (PCA) of metabolomics dataset from aspartate carbamoyltransferase assay. (a) Heatmap crerated from dataset using R. (b) PCA analysis created from dataset using R. Samples were analysed in triplicates. Co1_C = Control samples cofactor mix 1, Co1_E = Treatment samples cofactor mix 1, Co2_C= Control samples cofactor mix 2, Co2_E= Traetment samples cofactor mix 2 and QC= pooled samples. Heatmap shows that one Cofactor 1 control sample seemed not to have been picked up proper as the majority of the metabolites seemed decreased compared to the other samples. This sample should be discarded for analysis. No clear separation is apparent between the two groups. Heatmap colours are represented in a gradient from dark blue (metabolite not present or significantly decreased, green (decreased, not significant), yellow (unchanged) and red (significantly increased).

To establish if this protein is essential for bloodstream form trypanosomes an aspartate carbamoyltransferase^{RNAi} cell line (G197^{RNAi}) was created and the transcription of the gene knocked down by RNA interference. Induction of knock down with 1 µg ml^{-1} tetracycline showed no change in growth phenotype, but induction with and 5 µg ml^{-1} tetracycline did (see Figure 27 (a)). Reverse transcription with RT-PCR showed a knockdown effect of about 70% (Figure 3.27 (b)).



(a)



(b)

Figure 3.27.: Growth curve of G197^{RNAi} cell line (a) and relative RNA abundance in G197^{RNAi} cell line, as determined by rt RT-PCR (b)
(a) Culture of G197^{RNAi} , induced every 24 h with 5 µg ml^{-1} tet (+) and without tet (-) , was monitored for five days for differences in growth rate. Cell flasks were set up in triplicates.
(b) Two cultures of G197^{RNAi} were set up, one was induced with 5 µg ml^{-1} tetracycline (+tet) every 24 hours. Samples were taken (5 x 10^{7} cells) at timepoint 48h, 72h and 96h and RNA extracted and relative RNA abundance of transcript of interest was assessed by rt RT-PCR (in triplicate). Timepoints were combined for graph as there was no change in RNA abundance between different timepoints. Protein GPI-8 was used as standard.

To establish the function of this protein knockdown was induced in G197^{RNAi} and after 72 hours metabolites were extracted from uninduced and induced cells and analysed on LC-MS. The most significant changes in this dataset were L-glutamate (increasing levels in RNAi line) and UDP and uracil (decreasing levels in RNAi line). The pathway involving aspartate carbamoyltransferase and metabolite levels are shown in Figure 3.28. This pathway has been shown to be active in bloodstream form trypanosomes and a recently performed study with U- ^{13}C D-glucose showed labelling of L-aspartate from D-glucose and the downstream metabolites of the pyrimidine biosynthesis (Creek et al., 2015).

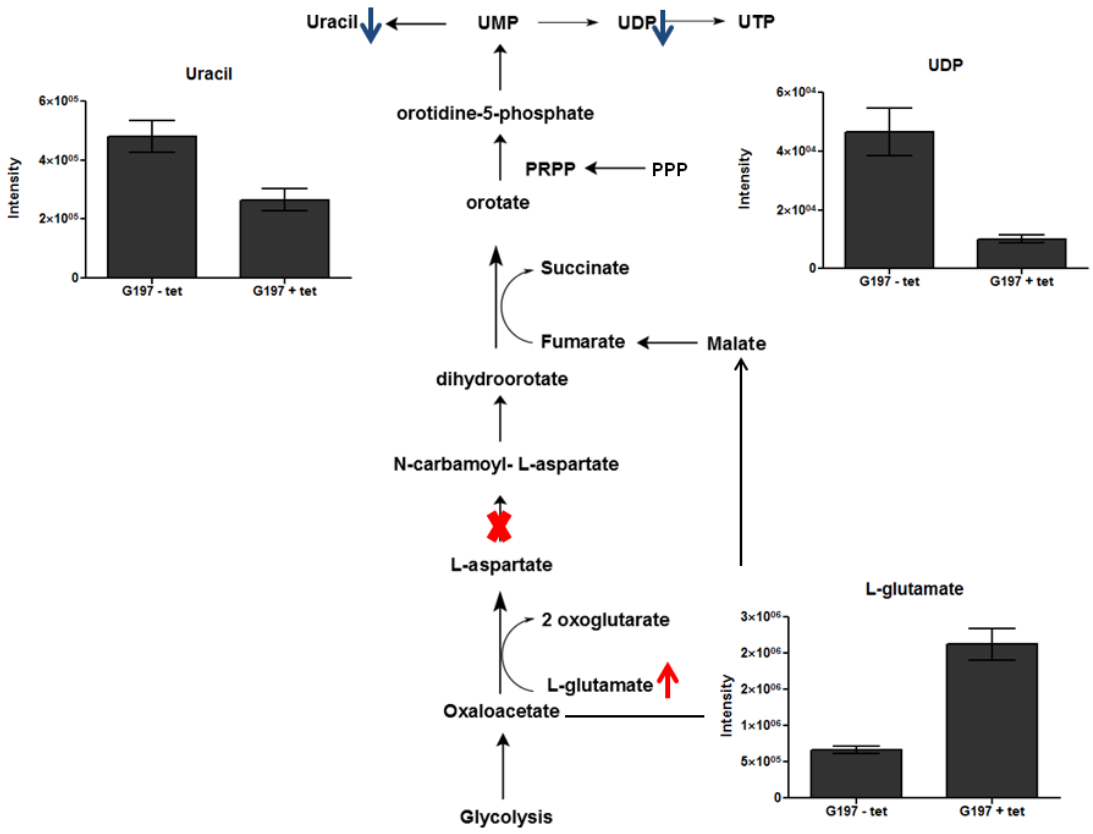


Figure 3.28: Pyrimidine biosynthesis pathway, adapted from metacyc.org. Red cross indicates reaction catalysed by aspartate carbamoyltransferase. Metabolites changing significantly in knockdown line are shown. L-glutamate increases in knockdown line, while uracil and UDP decrease. Abbreviations: PPP: Pentose phosphate pathway, PRPP: Phosphoribosyl pyrophosphate, UDP: Uridine diphosphate, UMP: Uridine monophosphate.

3.2.3.8. Arginase / Agamtinase-like protein, putative (G131)

Vincent *et al*. (2012) discovered that the gene Tb927.8.2020, putatively annotated as 'arginase' in TritrypDB, has no arginase activity. The annotation was later changed to agmatinase-like protein, however, several untargeted metabolite profiling approaches have failed to show this (Figure 3.29).

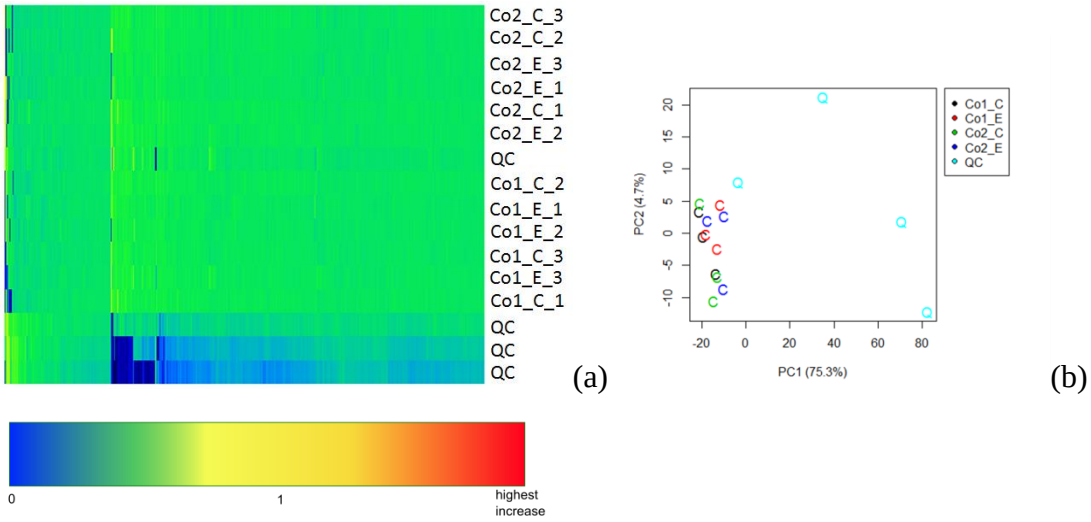


Figure 3.29: Heatmap and principal component analysis (PCA) of metabolomics dataset from arginase assay. (a) Heatmap of untargeted metabolite profiling approach with yeast extract and two cofactor mixes (3 replicates). The visible changes occur in the quality control (QC) samples, which are run at the beginning, in the middle and at the end of the run, with decreasing detected metabolites from QC1 to QC 4. Those changes are most likely due to sample degradation. (b) PCA analysis of dataset. Except for QC samples, no clear differences between the sample groups. Heatmap colours are represented in a gradient from dark blue (metabolite not present or significantly decreased, green (decreased, not significant), yellow (unchanged) and red (significantly increased).

To investigate the possibility of metal-ion dependency of Tb927.8.2020, heterologously expressed protein from *E. coli* Rosetta (DE3) pLysS was purified and incubated with yeast extract and three additional metal-ions (10 µM: manganese (Mn^{2+}), cobalt (Co^{2+}) and zinc (Zn^{2+})). Although the availability of Mn^{2+} and Co^{2+} is very low in trypanosomes, it has been shown in the case of phosphoglycerate mutase that replacement of a native metal by an alternative can increase the activity of a metalloenzyme (Fuad *et al.*, 2011). A heatmap of the created dataset of 'arginase' with added metal-ions (Figure 3.30) clearly shows changes between control and treatment samples. In total, about 20% of metabolites detected showed changes in levels between control and treatment samples.

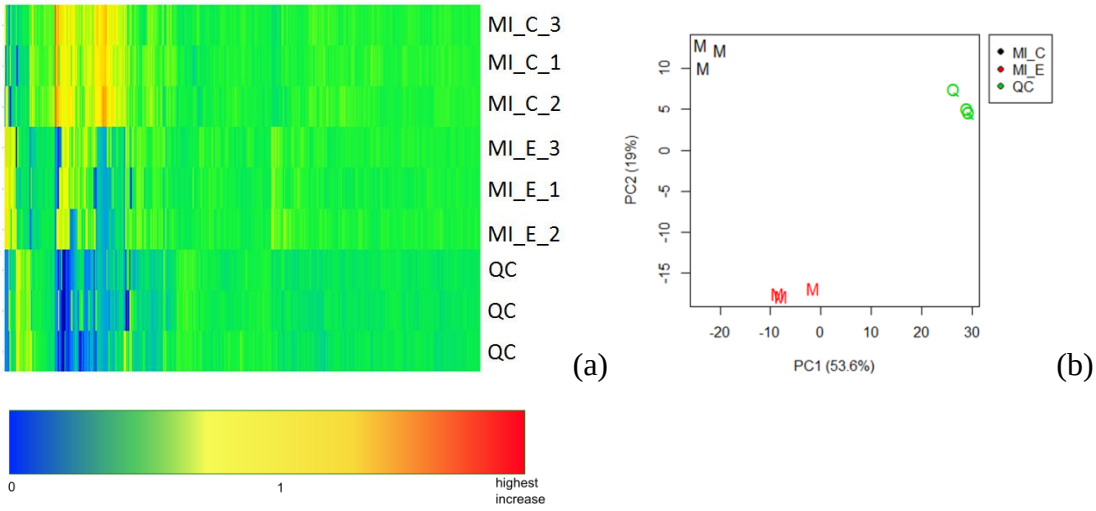


Figure 3.30: Heatmap and principal component analysis (PCA) of metabolomics dataset from arginase assay with metal-ions added. (a) Heatmap of untargeted metabolite profiling approach with yeast extract and a mix of three metal ions (3 replicates). Metal-ions used were Co2+, Mn2+ and Zn2+. Changes detected in the quality control (QC) samples are probably due to sample degradation. However, Samples treated with enzymes show decreased levels in metabolites, mainly in amino acid metabolism. (b) PCA analysis of dataset shows a clear separation between control, treatment and QC. Heatmap colours are represented in a gradient from dark blue (metabolite not present or significantly decreased, green (decreased, not significant), yellow (unchanged) and red (significantly increased).

According to putative metabolite annotation, a high proportion of changes taking place in the amino acid metabolism and a few biochemical reactions could be plausible:

- L-histidine + NADH → L-histidinal + NAD, catalysed by histidinal dehydrogenase (Figure 3.31).

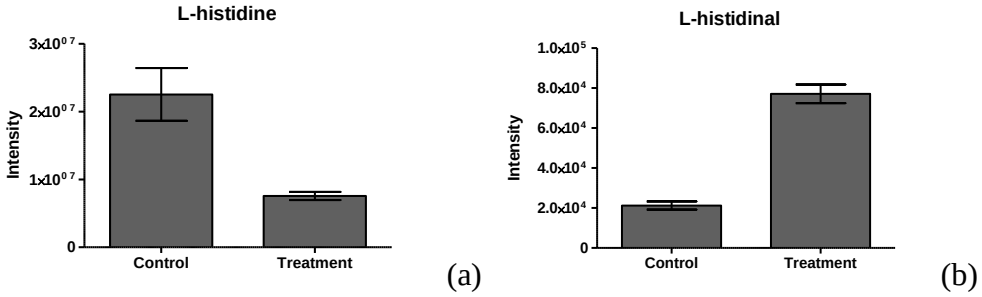


Figure 3.31: Significant changes in metal-ion 'arginase' dataset. (a) L-histidine and (b) L-histidinal.

However, L-histidinal seems to be a fragment of the dipeptide N-glycyl L-leucine and has been annotated as a possible fragment in IDEOM.

- Imidazole-5-pyruvate

Imidazole-5-pyruvate is a metabolite within the L-histidine and imidazole-lactate degradation pathway.

Reaction involving this metabolite (as taken from metacyc.org):

(1) L-histidine + 2-oxoglutarate ↔ L-glutamate + imidazole-5-pyruvate

(2) imidazole-lactate + NAD(P)$^+$ ↔ imidazole-5-pyruvate + NAD(P)H

(3) L-histidine + pyruvate ↔ imidazole-5-pyruvate + L-alanine

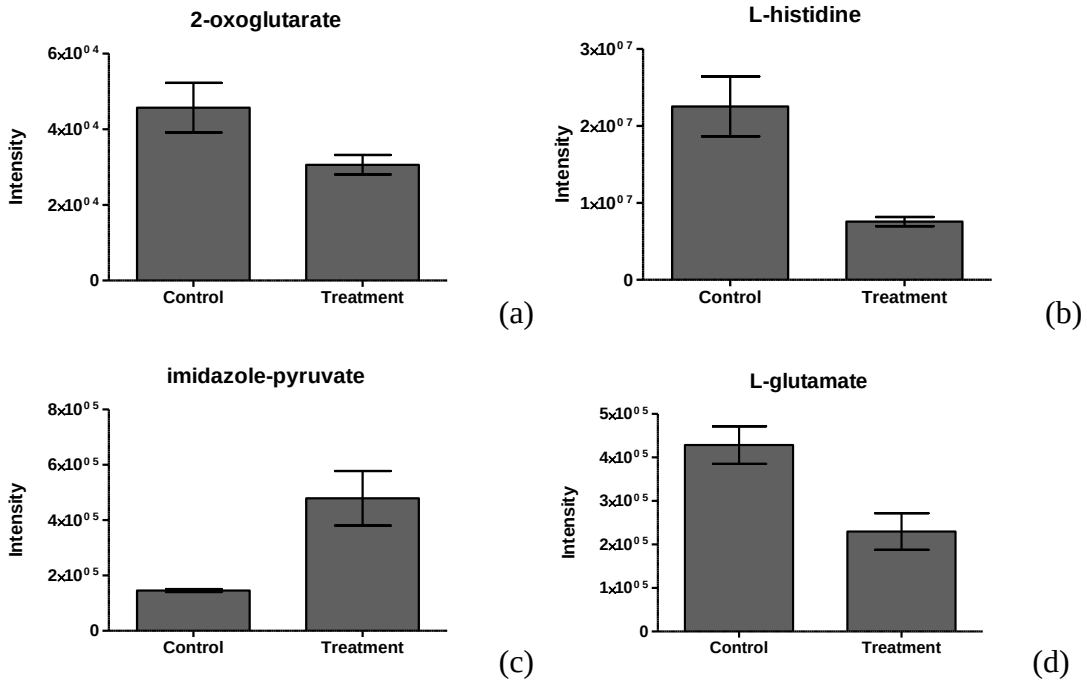


Figure 3.32: Significant changes in metal-ion 'arginase' dataset. (a) 2-oxoglutarate, (b) L-hsitidine, (c) imidazole-pyruvate and (d) L-glutamate.

A peak annotated as imidazole-5-pyruvate shows increasing peak intensity in the enzyme treated sample. But imidazole-5-pyruvate elutes within the basepeak of S-methyl-1-thiol-D-glycerate and has therefore been annotated as a fragment of that metabolite (Figure 3.32). Furthermore, the reaction involving imidazole-5-pyruvate does not appear to happen in this dataset. Imidazole-lactate was not detected, but listed as an isomer of 4-Imidazolone-5-propanoate, but again this metabolite seems to be a fragment of Adenosine. Also, levels of imidazole-lactate increases in enzyme treated sample set. Pyruvate had 2 isomers, one increasing slightly, the other one decreasing slightly, both possible fragments of another metabolite. L-alanine was not detected.

- (S)-1-Pyrroline-5-carboxylate

(S)-1-Pyrroline-5-carboxylate could be produced from L-glutamate in an NADH or NADPH dependent reaction. No cofactors were added to this sample set, however, it is possible that NADH or NADPH was present in the yeast extract. But, similar to imidazole-5-pyruvate, it elutes within the basepeak of S-methyl-1-thiol-D-glycerate (Figure 3.33 and 3.34).

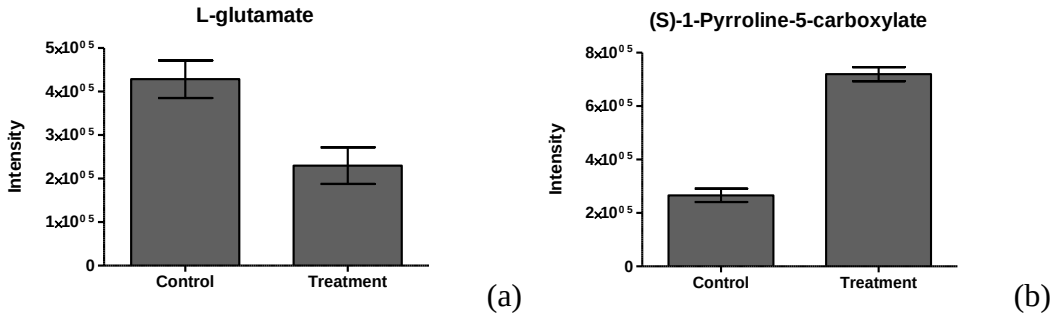


Figure 3.33: Average peak intensities of metabolites (S)-1-Pyrroline-5-carboxylate (a) and L-glutamate (b).

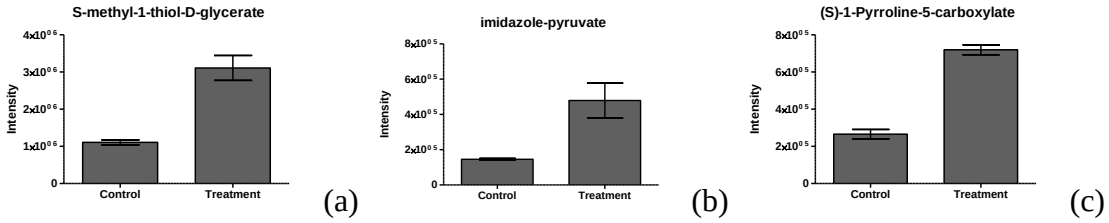


Figure 3.34: Detected basepeak S-methyl-1-thiol-glycerate (a) with the two potential fragments, detected as imidazole-pyruvate (b) and (c) (s)-1-pyrroline-5-carboxylate.

Although the function of the enzyme could not be determined from this experiment, the protein annotated as 'arginase' appears to decrease the levels of several amino acids. This could have been an indication of the enzyme being a non-specific deaminase. However, the corresponding keto acids were not detected or showed no differences in control and treatment samples. In the next step to determine the function of this enzyme, the 'arginase' was incubated in yeast extract, spiked with uniformly labelled ^{13}C-L-methionine and in a reaction mix with metal-ions and u-^{13}C-labelled L-methionine on its own. L-methionine was chosen as it was one amino acid decreasing the most in the previous dataset. The results are shown in Figure 3.35, but as previous results it only shows the decreasing level of the amino acids (in this case L-methionine), but no labelled product was detected. There could be several explanations for this, (1) ZIC-pHILIC is not a suitable column for this compound. (2) the levels of product are too low to be detected as a peak. As seen in Figure 3.31. (b) the detected labelled 13C L-methionine was quite low due to the high levels of unlabelled L-methionine in the yeast extract. (3) product does not ionise well and was therefore not detected. All metabolites significantly changed in the metal-ion 'Arginase' dataset are listed in Table 3.6.

Mass	RT (in mins)	Formula	Isomers	Putative metabolite identification	Treatment	Control
160.085	7.433	C6H12N2O3	4	D-Alanyl-D-alanine	**9.44**	1.00
139.075	7.361	C6H9N3O	3	L-Histidinal	**3.63**	1.00
195.053	7.601	C9H9NO4	9	Dopaquinone	**3.32**	1.00
154.038	7.690	C6H6N2O3	2	Imidazol-5-yl-pyruvate	**3.28**	1.00
113.048	7.703	C5H7NO2	6	(S)-1-Pyrroline-5-carboxylate	**2.71**	1.00
145.074	7.667	C6H11NO3	9	6-Amino-2-oxohexanoate	**2.28**	1.00
72.021	7.857	C3H4O2	4	Methylglyoxal	**1.27**	1.00
117.079	8.948	C5H11NO2	16	Betaine	**1.19**	1.00
193.074	4.919	C10H11NO3	10	Phenylacetylglycine	**0.88**	1.00
179.058	6.883	C9H9NO3	6	Hippurate	**0.84**	1.00
169.085	9.440	C7H11N3O2	5	N(pi)-Methyl-L-histidine	**0.76**	1.00
188.116	6.734	C8H16N2O3	7	N6-Acetyl-L-lysine	**0.73**	1.00
131.058	10.391	C5H9NO3	14	trans-4-Hydroxy-L-proline	**0.73**	1.00
103.063	10.800	C4H9NO2	14	4-Aminobutanoate	**0.72**	1.00
165.079	8.362	C9H11NO2	7	L-Phenylalanine	**0.71**	1.00
220.085	8.672	C11H12N2O3	3	5-Hydroxy-L-tryptophan	**0.70**	1.00
142.074	9.817	C6H10N2O2	1	Ectoine	**0.69**	1.00
160.048	8.965	C5H8N2O4	2	N-Formimino-L-aspartate	**0.69**	1.00
116.047	4.629	C5H8O3	9	5-Oxopentanoate	**0.66**	1.00
216.111	8.096	C9H16N2O4	3	γ-Glutamyl-γ-aminobutyraldehyde	**0.65**	1.00
227.079	8.357	C10H13NO5	1	L-Arogenate	**0.65**	1.00
218.127	6.826	C9H18N2O4	5	N2-(D-1-Carboxyethyl)-L-lysine	**0.64**	1.00
145.085	10.887	C5H11N3O2	3	4-Guanidinobutanoate	**0.62**	1.00
133.037	10.072	C4H7NO4	4	L-Aspartate	**0.59**	1.00
130.063	4.274	C6H10O3	17	4-Methyl-2-oxopentanoate	**0.58**	1.00
204.111	7.497	C8H16N2O4	5	N6-Acetyl-N6-hydroxy-L-lysine	**0.56**	1.00
147.053	9.914	C5H9NO4	14	L-Glutamate	**0.54**	1.00
200.977	11.135	C3H7NO5S2	1	S-Sulfo-L-cysteine	**0.50**	1.00
226.107	8.227	C9H14N4O3	3	Carnosine	**0.49**	1.00
181.074	9.880	C9H11NO3	11	L-Tyrosine	**0.46**	1.00
119.058	10.344	C4H9NO3	11	L-Threonine	**0.46**	1.00
165.046	9.620	C5H11NO3S	4	L-Methionine S-oxide	**0.45**	1.00
129.043	10.173	C5H7NO3	6	1-Pyrroline-4-hydroxy-2-carboxylate	**0.44**	1.00
75.032	10.980	C2H5NO2	3	Glycine	**0.44**	1.00
175.048	9.676	C6H9NO5	4	N-Acetyl-L-aspartate	**0.42**	1.00
146.069	10.575	C5H10N2O3	6	L-Glutamine	**0.38**	1.00
164.047	4.251	C9H8O3	13	Phenylpyruvate	**0.36**	1.00
131.095	8.742	C6H13NO2	12	L-Leucine	**0.36**	1.00
155.069	10.478	C6H9N3O2	5	L-Histidine	**0.34**	1.00

132.053	10.669	C4H8N2O3	6	L-Asparagine	**0.32**	1.00
105.043	10.983	C3H7NO3	3	L-Serine	**0.31**	1.00
87.032	10.535	C3H5NO2	3	2-Aminoacrylate	**0.31**	1.00
149.051	9.102	C5H11NO2S	5	L-Methionine	**0.28**	1.00
174.112	18.361	C6H14N4O2	2	L-Arginine	**0.27**	1.00
175.096	10.983	C6H13N3O3	3	L-Citrulline	**0.26**	1.00
246.133	10.931	C9H18N4O4	2	N2-(D-1-Carboxyethyl)-L-arginine	**0.23**	1.00
132.090	10.962	C5H12N2O2	6	(2R,4S)-2,4-Diaminopentanoate	**0.17**	1.00

Table 3.6: Significant changes detected in amino acid metabolism in MI sample set. Control samples had purified recombinant enzyme added after reaction was quenched, while the treatment had the enzyme added for 30min incubation. Metabolites identified with confidence 10 (matching authentic standard) are highlighted in grey.

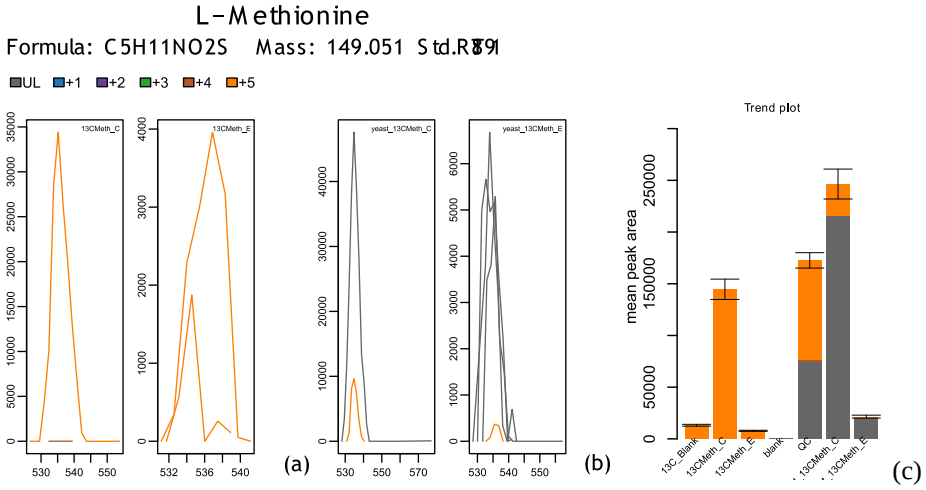


Figure 3.35: In vitro assay with ^{13}C-labelled L-methionine. (a) Uniformly ^{13}C-Methionine in 10 mM MOPS buffer + MI was incubated at 37°C for 30 min without (left) and with (right) 'Arginase'. (b) Yeast extract in 10 mM MOPS + MI was spiked with U-^{13}C-Methionine and incubated at 37°C for 30 minutes without (left) or with (right) 'Arginase'. (c) shows the trend plot of labelled L-methionine between the different sample sets.

Crystallography studies first suggested that the arginase-like protein is indeed a metalloenzyme and it was indicated that it binds significant quantities of Fe(II) ions, at least to a 1:1 ratio. The hypothesis is that Fe(II) could be liganded by a Try-His-Asp triad in the active site of the enzyme (Dr. David Christianson, University of Pennsylvania, personal correspondence). However, although changes in the dataset are evident using additional metal-ions as cofactors, it has recently been shown that the *T. brucei* 'arginase' does not bind metal-ions as previously thought (Hai et al., 2015). The observed binding of Fe(II) to the enzyme turned out to be weak binding to the HIS-tag of that protein. Why the changes, specifically to the amino acids, occur in the metabolomics dataset is unclear.

3.3. Discussion

The use of metabolomics for enzyme function identification is well documented, with this application used in multiple studies (de Carvalho et al., 2010; Larrouy-Maumus et al., 2013; Liscombe et al., 2010; Saito et al., 2006, 2009). However, in this project, I tried to apply this method to a variety of putatative enzymes to investigate its use in a high throughput approach for genome wide annotations and to establish a workflow to do so.
The focus was put on protein over expression, purification and assay optimisation. For cloning a ligase independent cloning system was used to cut down on time spent. Although, the use of Xa/LIC cloning system seemed initially more expensive, the time saved and the benefit of not having to use restriction enzymes made this approach more suitable. Protein over expression was performed in *E. coli* Rosetta (DE3) or Rosetta (DE3) pLysS, due to the speed and low cost of this system. As a lot of proteins could not be over expressed, for future work the use of trypanosomatids over expression systems is worth to be considered for proteins that cannot be obtained using *E. coli*. Protein purification was optimised by changing from Ni-NTA Spin Columns (Qiagen) to Poros MC20 column HPLC system. The quality of purified proteins was increased, which has to be considered as a vital part of this workflow. A few changes were also made to the assay as it was originally used in our lab (E. Kerhoven, thesis). The concentration of MOPS buffer was lowered from 40mM to 10mM, as MOPS was seen to block the ZIC-pHILIC column used (K. Burgess, Glasgow Polyomics). The change of concentration did not seem to affect the enzyme assays, although it was not directly tested, the results obtained from S-adenosylmethionine synthetase were performed in 10mM MOPS. A second extraction step was included, which increased the quality of the obtained dataset slightly. Data analysis was solely performed using IDEOM, as the obtained datasets were very complex, normally with around 1,000 putatively identified metabolites. The use of IDEOM showed to make those datasets manageable and easy to screen for changes in the dataset. The identified enzyme functions in this study, two commercial enzymes and a putative S-adenosylmethionine synthetase, showed the most significant changes in the metabolomics dataset straight away, although data still needed investigation, as the changes did not always immediately show the ‘correct’ metabolite identification.

Besides the S-adenosylmethionine synthetase, were the function could be shown, six additional enzymes were investigated. Four of them showed no significant changes in the datasets. For deoxyhypusine synthase it was recently shown that this enzyme’s activity is

regulated by a catalytically dead paralogue (Ngyuen et al., 2013), which might explain why no changes were seen.

Deoxyribose-phosphate aldolase and aldo/keto reductase also showed no significant changes in the dataset. RNAi lines were created to further investigate their function, but the knock down achieved was only 10% and 40% respectively, so no metabolomics experiments were performed. The putative aspartate carbamoyltransferase showed no significant changes in the sample set with the *in vitro* assay approach. However, the RNAi line created showed growth defect by 70% knock down. Metabolomics analysis showed decreasing levels in UDP and uracil and increasing levels in L-glutamate (Figure 3.27) indicating that the enzyme is involved in the biosynthesis of pyrimidines.

The putative NAD synthase and citrate synthase need further investigation. Changes were seen in the obtained metabolomic datasets. However, a definite function could not be assigned to them.

The created workflow in this project took 7 days to complete for one protein, starting with PCR for cloning and finishing with samples for LC-MS analysis. Although two proteins could be prepared at the same time, the need for the Poros MC20 column HPLC system slows protein purification down as proteins cannot be purified at the same time (as was possible with the Ni-NTA Spin Columns). As purified proteins were used immediately after preparation two *in vitro* assays could be performed per week.

Chapter 4

4.1 Introduction

Metabolomics has proven to be an excellent tool for quantitative and qualitative analysis of low molecular weight metabolites and their interaction within a living cell (Dunn et al., 2011). Areas of research which have benefited from this approach include drug development, biomarker discovery and the exploration of new metabolic pathways. Whilst many metabolites can be easily identified, suggesting which pathways may be active, direct proof of this is missing. One possible method to circumvent this problem has been to combine an untargeted metabolomic approach with stable isotope labelling. This method has successfully been employed to study metabolic pathways in trypanosomes (Creek et al., 2012c); providing not only a snapshot of cellular metabolism, but also of direct pathway identifications by tracing labelled compounds.

Trypanosomes have different life stages and can adapt quickly to changes in their environment. Procyclic trypanosomes utilise L-proline instead of D–glucose as an energy source, as their environment in the mid-gut of the tsetse fly provides amino acids more readily (Bursell, 1963). The glucose metabolism has been very well studied in bloodstream form and procyclic trypanosomes and Creek et al (2015) have recently provided an extended form of glucose metabolism in bloodstream form trypanosomes by combining stable isotope labelling with metabolomics. It was shown that glucose enters many branches of metabolic pathways, including for example polyamine biosynthesis (via glucose labelled ATP, which enters the methionine cycle), succinate fermentation pathway (forming malate, fumerate and succinate) and nucleotide synthesis (via ribose phosphate).

In this chapter the metabolic pathways of L-methionine, L-proline and L-arginine are investigated using stable isoptope labelling combined with metabolomics to investigate their distribution within trypanosomes.

L-methionine was thought to be salvaged from methylthioadenosine (MTA) via the MTA cycle in trypanosomes (Berger et al., 1996); however, the absence of labelled L-methionine from glucose (Creek et al., 2015) indicates the lack of an active MTA cycle. Data obtained from U-^{13}C glucose showed 3-C labelled succinate and malate, which is consistent with their production via the succinate fermentation pathway. However, a significant amount of

malate and succinate were not labelled from glucose. Further experiments using U-^{13}C L-glutamine showed the labelling of malate and succinate from L-glutamine, contributing to the intracellular presence of those metabolites. L-proline could also contribute to the production of malate and succinate, but so far the use of L-proline for metabolic purposes was only shown for procyclic trypanosomes.

Another well studied pathway is the biosynthesis of polyamines in trypanosome. L-ornithine and putrescine are important precursors leading to the production of trypanothione, the main thiol in trypanosomes (Fairlamb et al., 1985). In fact, inhibition of the enzymes ornithine decarboxylase, which catalyses the conversion of L-ornithine to putrescine, by the drug eflornithine, leads to cell death (Fairlamb et al., 1989; Vincent et al., 2012). The classical route for biosynthesis of L-ornithine is via L-arginine and is catalysed by the enzyme arginase. Previous work has shown that arginase activity is absent in bloodstream form trypanosomes (Hai et al., 2014), but also indicated that L-ornithine can still be produced from L-arginine (I. Vincent, thesis). To investigate the L-arginine metabolism in bloodstream form trypanosomes U-^{13}C L-arginine was used.

4.1.1. L-methionine

The sulfur containing amino acids L-methionine and L-cysteine play an important role in protein synthesis, methylation processes in the cell (L-methionine only) and biosynthesis of polyamines and glutathione. The importance and role of these two compounds has been widely discussed in several reviews for trypanosomes and other organisms (Nozaki et al., 2005; Stipanuk, 2004; Walker and Barrett, 1997; Willert and Phillips, 2012). Not only does L-methionine play a vital role in cell survival, but the polyamine pathway has also been in research focus, due to the unique features it possesses in trypanosomes that can be useful for drug development against trypanosomiasis. Eflornithine, one of the drugs currently in use against this deadly disease, is known to inhibit the biosynthesis of polyamine pathway and causes cell death of the parasites (Yarlett and Bacchi, 1988; Vincent et al., 2012). In this chapter, I explore the metabolism of L-methionine in *T. brucei*, using U-^{13}C L-methionine and metabolomics.

4.1.1.1. Protein synthesis

L-methionine is the starting amino acid in protein synthesis. It also plays a vital role in the active site of proteins as, due to its sulphur atom, it is easily oxidised (L-methionine to methionine sulfoxide), but does not seem to have the same importance to structure and stability as L-cysteine, whose highly reactive thiol-group has a big impact on structure and stability of proteins (Nozaki et al, 2005).

4.1.1.2. Methylation processes and methionine cycle

The L-methionine intermediate S-adenosylmethionine is an important methyl donor in many methylation processes of the cell. Binding of an adenosyl-group from ATP to the sulphur in L-methionine puts a positive charge to the sulphur and activates the methyl-group to be reactive with other compounds. It is estimated that about 95% of the acquired S-adenosylmethionine pool gets used for methylation processes in the cells, at least in mammals (Walker and Barrett, 1997). For trypanosomes the estimated figure is 90% (Bacchi and Yarlett, 1993). The resulting S-adenosylhomocysteine is toxic to cells, so is quickly being converted to L-homocysteine (see Figure 4.1, reaction 3). L-homocysteine can be converted back to L-methionine (see Figure 4.1, reaction 4) and therefore concluding the methionine cycle. An alternative route for L-homocysteine is the transulfuration pathway, where it is being converted to cystathionine and further to L-cysteine (Figure 4.1, reaction 5 and 8).

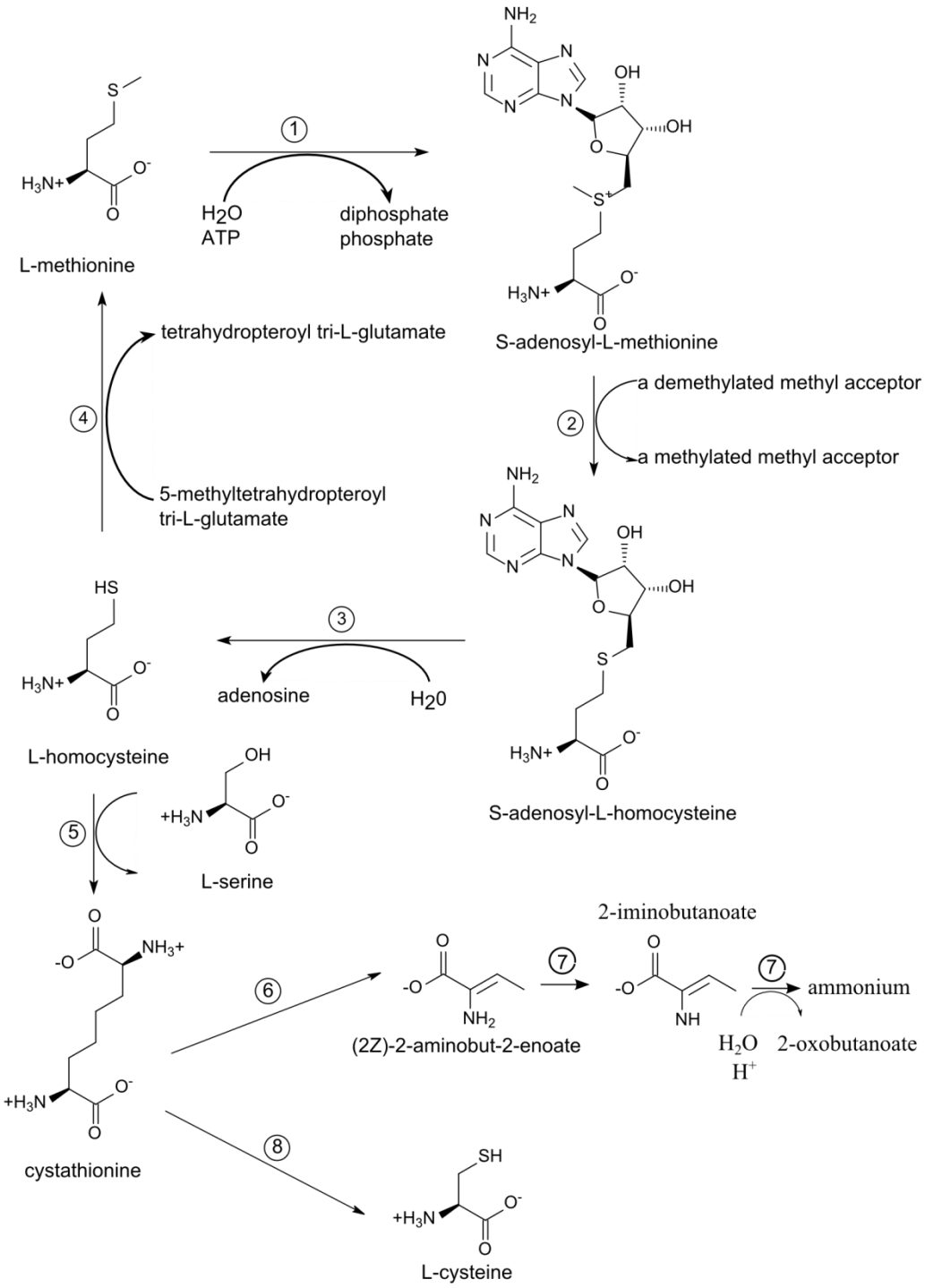


Figure 4.1: Methionine cycle as adapted from metacyc.org. This represents the parasitic protozoan pathway as described by Walker and Barrett (1997) Key to enzymes: 1- S-adenosylmethionine synthetase; 2- various methyltransferase; 3- S-adenosylhomo-cysteine hydrolase; 4- 5'-Methyltetra-hydrofolate:homocysteine methyltransferase (methionine synthase). For reaction 4 a different enzyme is indicated for mammalian host (Betaine:homocysteine methyltransferase, EC 2.1.1.5). Reaction involves L-homocysteine and betaine to form L-methionine and dimethylglycine. This enzyme is supposedly absent in trypanosomes. 5- Cystathione β-synthase (EC 4.2.1.22), 6/8-γ-Cystathionase (EC 4.4.1.1), 7- spontaneous reaction

In *T. brucei* strain 927, a gene annotated as encoding for S-adenosylmethionine synthase exists in multiple copies. S-adenosylmethionine synthase activity has been shown to be

present in trypanosome extracts (Bacchi and Yarlett, 1993) and results from an in vitro assay using metabolite profiling on a recombinant S-adenosylmethionine synthase expressed from *E. coli* confirms the presence of this enzyme in trypanosomes (see chapter 3.2.3.1). When S-adenosylmethionine gets converted to S-adenosylhomocysteine various methyltransferases play a role in this reaction. A search in the trypanosome database TritrypDP (version 8.1) reveals 41 genes annotated as methyltransferases, some of them putative. S-adenosylhomocysteine to L-homocysteine is being catalysed by S-adenosylhomocysteine hydrolase (or adenosylhomocysteinase (EC 3.3.1.1)). The trypanosome genome contains one gene that has been putatively identified as S-adenosylhomocysteine hydrolase (Tb927.11.9590). This is the second branchpoint in the methionine cycle, as L-homocysteine can be converted back to L-methionine or into L-cystathionine. To convert L-homocysteine back to L-methionine, two genes have been putatively identified in *T. brucei* strain 927 to catalyse this reaction.

(1) Homocysteine S-methyltransferase (EC 2.1.1.10, Tb927.1.1270). This enzyme transfers a methyl-group from S-adenosylmethionine to L-homocysteine.

(2) 5'-Methyltetra-hydrofolate:homocysteine methyltransferase (EC 2.1.1.14, Tb927.8.2610), as shown in Figure 4.1.) transfers a methyl-group from 5'-methyltetrahydropteroyl tri-L-glutamate to form L-methionine from L-homocysteine.

L-homocysteine can be converted to cystathione by cystathionine β-synthase. This pathway belongs to the biosynthesis of L-cysteine, where L-cysteine can be produced from L-cystathione by γ-cystathionase.

4.1.1.3. Biosynthesis of polyamines

S-adenosylmethionine can be converted into S-adenosyl-5'-deoxy-3',5'-methylthio-adenosine (decarboxylated S-adenosylmethionine, dSAM) by S-adenosylmethionine decarboxylase (AdoMet decarboxylase) (Figure 4.2). This is where L-methionine enters the biosynthesis of polyamines. In comparison to the mammalian AdoMet decarboxylase, the trypanosome enzyme is only weakly activated by putrescine (Bitonti et al., 1986), but it is activated by dimerization with an inactive paralogue (Prozyme) of the active AdoMet decarboxylase (Pegg, 2009; Willert and Phillips, 2012). Decarboxylated S-adenosylmethionine gets converted to 5-methylthio-adenosine and spermidine by spermidine synthase (Figure 4.3)

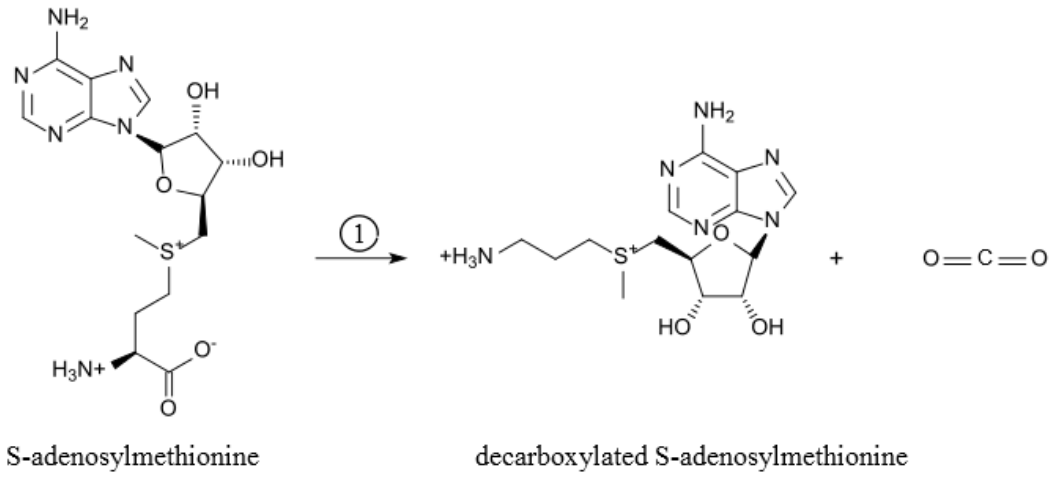


Figure 4.2: Methionine entering the biosynthesis of polyamines. S-adenosylmethionine gets converted to decarboxylated S-adenosylmethionine by SAM decarboxylase (1)

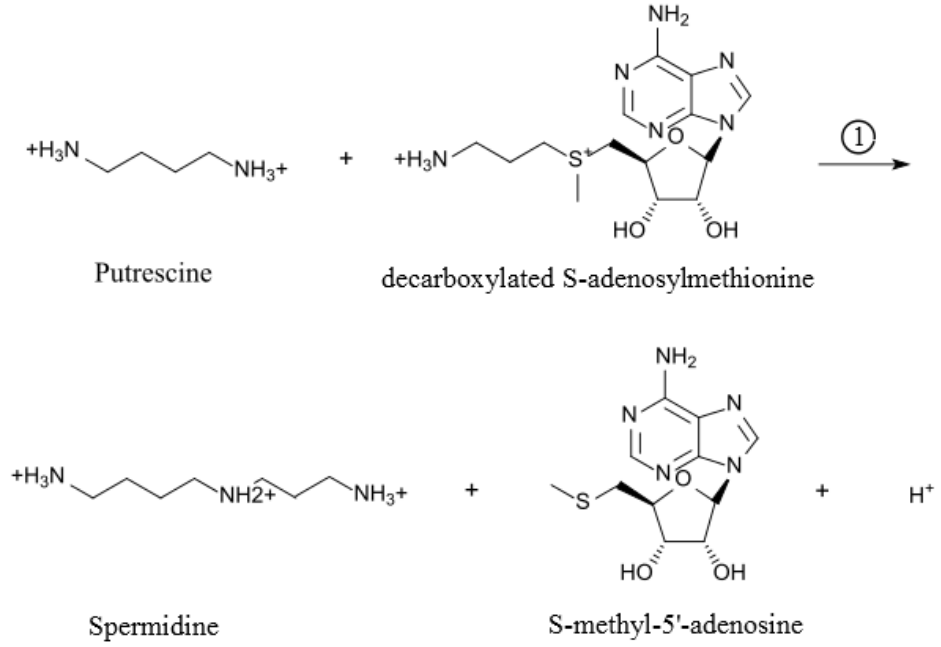


Figure 4.3: dSAM and putrescine get converted to spermidine and S-methyl-5'-adenosine by Spermidine synthase (EC 2.5.1.16) (1). Spermidine is being utilised further in the synthesis of polyamines, while S-methyl-5'-adenosine is entering the L-methionine salvage pathway (MTA or Yang cycle).

Spermidine is being further metabolised to the trypanosomatid specific trypanothione (Fairlamb, 1989), which is essential to those parasites by protecting them from oxidative stress. As seen in Figure 4.4, spermidine and glutathione form glutathionylspermidine, this reaction is catalysed by glutathionylspermidine synthase (EC 6.3.1.8).

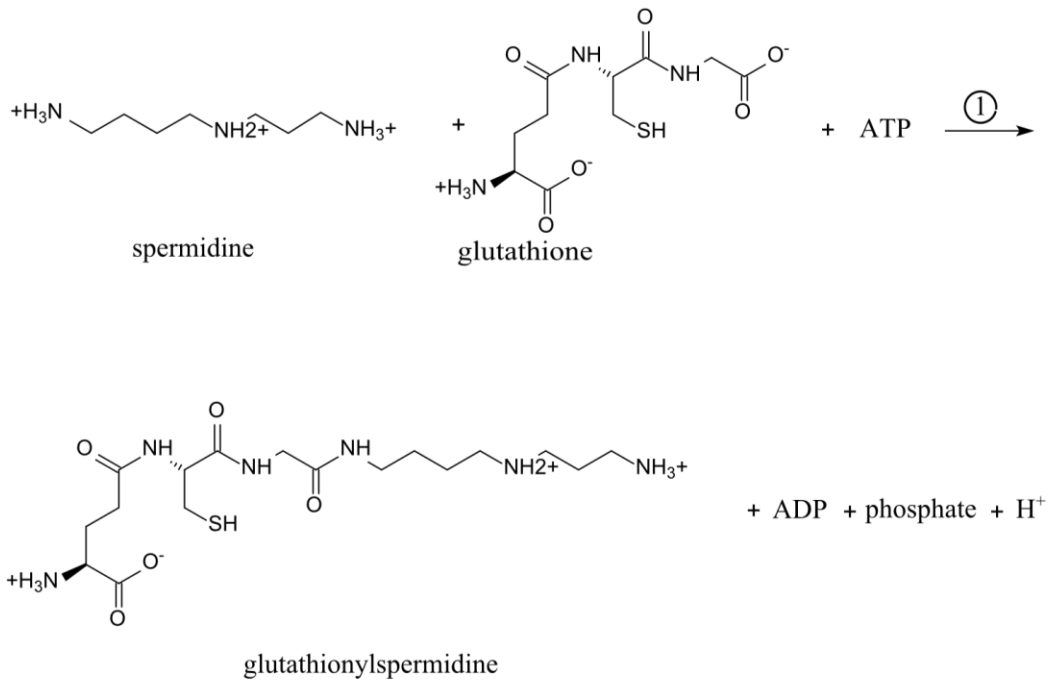


Figure 4.4: Formation of glutathionylspermidine from spermidine and glutathione, as adapted from metacyc.org. 1- Glutathionylspermidine synthetase (EC 6.3.1.8)

Trypanothione is formed by adding an additional glutathione to glutathionylspermidine, catalysed by trypanothione synthase (EC 6.3.1.9) as seen in Figure 4.5. Both trypanothione synthase and glutathionylspermidine synthase require ATP.

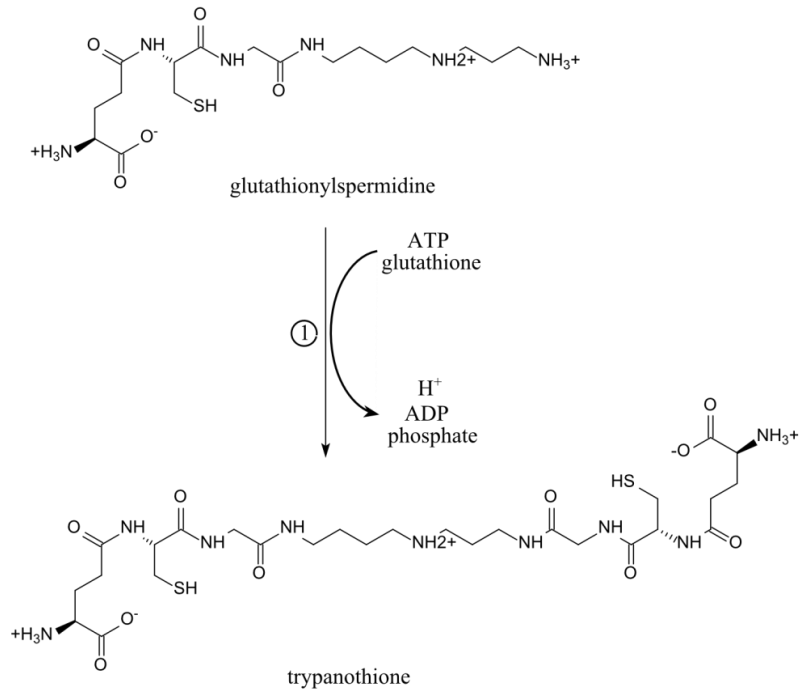


Figure 4.5: Last step in biosynthesis of trypanothione. Glutathionylspermidine and glutathione form trypanothione, reaction adpted from metacyc.org. 1- Trypanothione synthase (EC 6.3.1.9).

4.1.1.5. Methiolthioadenosine or Yang cycle

The methylthioadenosine (MTA) or Yang cycle is an L-methionine salvage pathway. Over seven steps the cell toxic product methylthioadenosine can be converted back to L-methionine (see Figure 4.6). Aspartate aminotransferase has been identified as the amino group donor of the final step in this pathway (2-oxo-4-methylthiobutyrate to L-methionine) for trypanosomes (Berger et al., 2001).

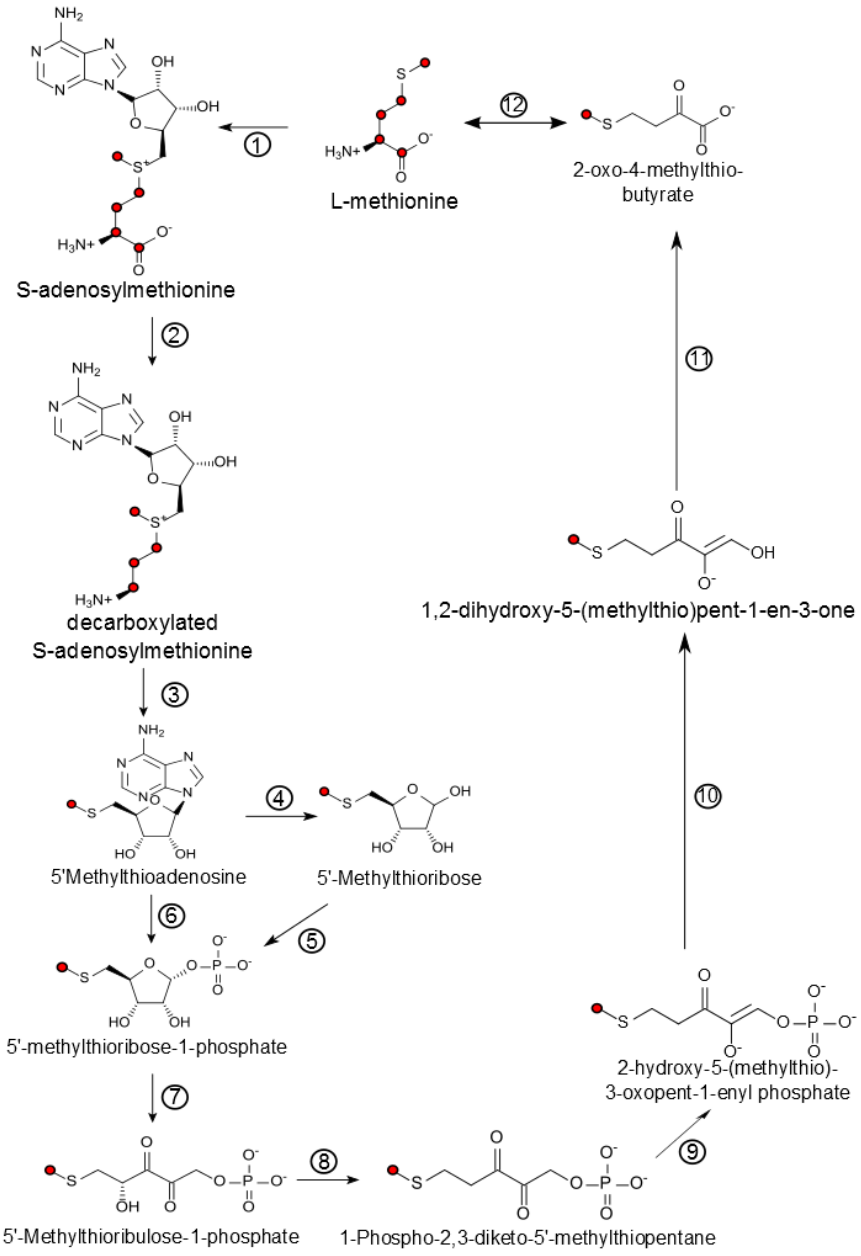


Figure 4.6: methylthioadenosine (MTA) or Yang cycle as adapted from metacyc.org. Enzyme key: **1**-S-adenosylmethionine synthetase, **2**-SAM decarboxylase, **3**-spermidine synthase, **4**-methylthioadenosine nucleosidase, **5**- methylthioribose kinase, **6**- methylthioadenosine phosphorylase, **7**- 5-methylthioribose-1-phosphate isomerase, **8**- 5-methylthioribulose-1-phosphate dehydratase, **9**- 2,3-diketo-5-methylthio-1-phosphopentane enolase, **10**-2-hydroxy-3-keto-5-methylthio-phosphopentene phosphatase, **11**- acireductone dioxygenase, **12** – methionine oxo-acid transaminase (metacyc). Expected carbon contribution from L-methionine throughout this cycle is indicated by a red dot.

4.1.2. L-arginine metabolism

L-arginine is reportedly involved in several metabolic pathways in trypanosomatids (Pereira, 2014).

- Urea cycle

  L-arginine gives an amidino group to form urea. Catalysed by the enzyme arginase, L-arginine is converted to L-ornithine and urea. Arginase is present in two isoforms in most mammals. Arginase I is primarily located in the liver and takes part in the urea cycle, while Arginase II is located in the mitochondria and regulates L-arginine and L-ornithine concentration within the cell (Morris, 2002). Arginase is known to be present in Leishmania parasites, but does not seem to be present in *T. brucei.*

- Polyamine biosynthesis

  L-ornithine is converted to putrescine which is the starting diamine in the biosynthesis of the polyamines spermidine and trypanothione (Fairlamb et al., 1985; Tabor and Tabor, 1984).

- Alternative pathway for L-proline biosynthesis

  L-proline can be synthesised via L-ornithine and L-glutamate 5-semialdehyde (Hird et al., 1983)

- Phosphagen synthesis

  L-arginine transfers an amidino group to an amino acceptor, forming guanidine derivates (Hird, 1986). Those phosphagens can be used for energy storage. Studies on *T. cruzi* and *T. brucei* have shown that arginine phosphates are important for energy storage and at least in the case of *T. cruzi* can protect the parasite against oxidative stress (Pereira, 2014)

- Nitric oxide synthesis

  L-arginine is the precursor for the production of nitric oxide by nitric oxide synthase (NOS). NOS and arginase compete for L-arginine pools.

4.1.2.1. Urea cycle and polyamine synthesis from L-arginine

Previous studies from I. Vincent (University of Glasgow, thesis) have shown that arginase, the classical enzyme to produce L-ornithine from L-arginine, seems to be absent from *T. brucei*. Metabolomic studies using recombinant ‘arginase’ from *T. brucei* (Chapter 3) and an arginase knockout line (Hai et al., 2014) has shown no arginase activity.

Although it was evident that L-ornithine could be produced from L-arginine, it is unlikely that L-ornithine biosynthesis happens via arginase as no urea could be detected in bloodstream form trypanosomes cell extract using a commercial arginase detection kit (Vincent, thesis). That would lead to the conclusion that Ornithine is produced via a different pathway than the arginase route, but speculation that the urea cycle could be operative in bloodstream form trypanosomes in reverse direction (Arginine – Citrulline – Ornithine) could not be proven.

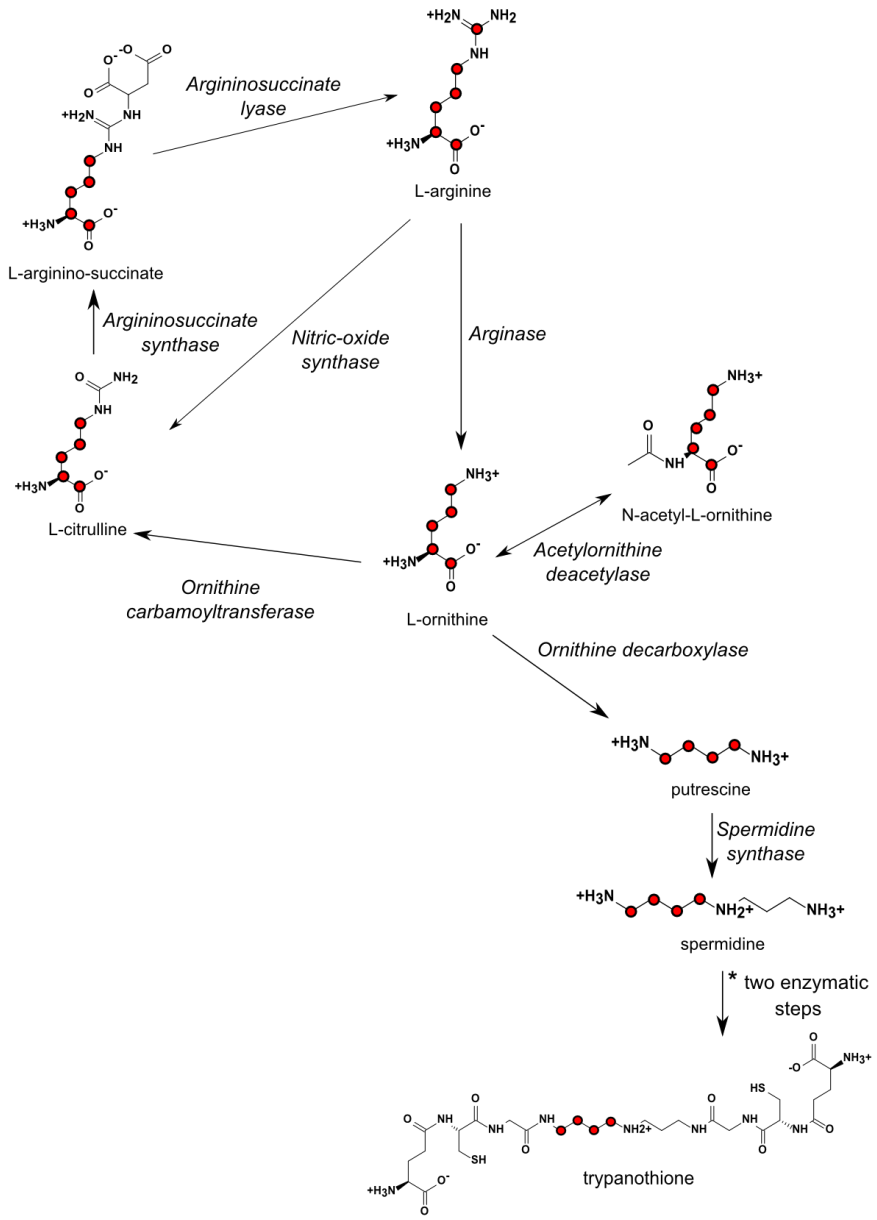


Figure 4.7: Predicted metabolites involved in the biosynthesis and degradation of L-ornithine. Carbons transferred between metabolites, originating from L-arginine, are highlighted in red. Enzymes involved in those reactions are shown in italic. * Two enzymatic steps are shown in Figure 4.4 and 4.5.

4.1.2.2. Phosphagens from L-arginine

Phosphoarginine has been shown to be an important phosphagen in many organisms and has been studied in *T. cruzi* where it has shown to protect the cells from oxidative stress (Miranda et al., 2006). *T. brucei* genome encodes for three arginine kinases, TbAK1-3, each of the isoforms is found in a specific subcellular compartment. TbAK1 is only present in the flagellum, while TbAK2 and TbAK3 in the glycosome and cytosol respectively (Voncken et al., 2013). When oxidative stress is induced in *T. cruzi*, levels of Phosphoarginine increase (Pereira, 2014). However, when a similar experiment was performed in *T. brucei*, this was not confirmed and arginine kinase knock out lines showed that they are not essential in *T. brucei* (D H Kim, unpublished data).

4.1.3. L-proline

In procyclic trypanosomes L-proline can be utilised as main carbon source for the energy metabolism when glucose supply is limited in the insect host (Bursell, 1963). It has been shown that procyclics can switch their energy metabolism regarding ATP production very quickly between substrate phosphorylation (D-glucose) and oxidative phosphorylation (L-proline) and reverse when nutrition supply is changed (Coustou et al., 2008). However, in bloodstream form trypanosomes L-proline has not been indicated for the use in energy metabolism, as parasites in the bloodstream depend on the uptake of D-glucose and glycolysis for ATP production (Bringaud et al., 2006). The classical TCA cycle does not seem to be functional in trypanosomes, but part of this cycle seem to be operative nonetheless (van Weelden et al., 2005). L- Proline enters the TCA cycle in proline degradation (van Hellemond et al., 2005), with succinate and acetate being the endproducts (Van Weelden et al., 2005). Tracking ^{13}C labelled glucose in bloodstream form trypanosomes also revealed malate, succinate and fumarate being labelled (Creek et al, 2015). Malate, succinate and fumarate were also labelled from L-glutamine (DH Kim and F Achcar, unpublished), however, glucose and L-glutamine combined still did not account for the three compounds labelled. But could L-proline in bloodstream form trypanosomes also contribute to the production of succinate, fumarate and malate as in procyclics?

L-proline enters the cell and gets converted to (S) 1-pyrroline-5-carboxylate by proline dehydrogenase. (S) 1-pyrroline-5-carboxylate forms L-glutamate-5-semialdehyde in a spontaneous reaction and gets further converted to L-glutamate by 1-pyrroline-5-

carboxylase. L-glutamate enters the TCA cycle by being converted to 2-ketoglutarate, a reaction catalysed by glutamate dehydrogenase. 2-ketoglutarate is transformed to succinyl-CoA and further to succinate, fumarate and malate.

L-glutamate can also be involved in the biosynthesis of L-arginine, via N-acetyl-L-glutamate, N-acetyl-L-ornithine, L-ornithine, Citrulline, L-arginino-succinate and lastly L-arginine. However, the above mentioned dataset from trypanosome cultures incubated with ^{13}C-glutamate did not show any labelled L-arginine from L-glutamate. Vincent (thesis) also showed no labelling of L-arginine or L-ornithine occurring from L-proline when cells were incubated with ^{15}N-Proline. This concludes that L-proline does not seem to play a part in L-arginine or L-ornithine biosynthesis. To fully investigate if bloodstream form trypanosomes utilise L-proline for metabolic purposes, cells were cultured and incubated for 48h with ^{13}C-labelled L-proline.

4.2. Results

Metabolomics with uniformly ^{13}C-labelled compounds was used to study the metabolic pathways of L-methionine, L-arginine and L-proline in the bloodstream form *T. brucei*. Experimental design, as described in methods 2.4.2. and 2.4.3., involved adding 50 % U-^{13}C L-methionine (50µM final concentration) and 100% U-^{13}C L-arginine or L-proline (at a final convcentration of 200 µM) to the culture medium (CMM + 10% FBS gold). Cultured *T. brucei* strain 427 were incubated for 48 hours in the presence of the labelled compounds, with a starting concentration of 2 x 10^4 cells ml^{-1}. After 48 hours, cells were cooled down rapidly to 0°C by placing the falcon tube in dry ice / ethanol bath and cells were collected by slow centrifugation (to avoid cell lysis). Cell pellets were resuspended in extraction solvent (Chloroform Methanol Water, 1:3:1) and metabolites analysed by LC-MS (ZIC-pHILIC). Data analysis was achieved using mzMatch and mzMatch.ISO, using authentic standards and KEGG database for metabolite identification.

4.2.1. Global L-methionine metabolism

In total 17 labelled metabolites were detected in this dataset, belonging to the methionine cycle, L-cysteine and polyamine biosynthesis and the MTA or Yang cycle. Six metabolites seem to be a product of methylating processes from S-adenosyl-L-methionine and are shown in 4.2.6, the other metabolites were mapped into global metabolite map (Figure 4.8). Some intermediate metabolites belonging to the pathways were not detected, but were mapped nonetheless to give a better understanding of the L-methionine metabolism in the bloodstream form trypanosomes.

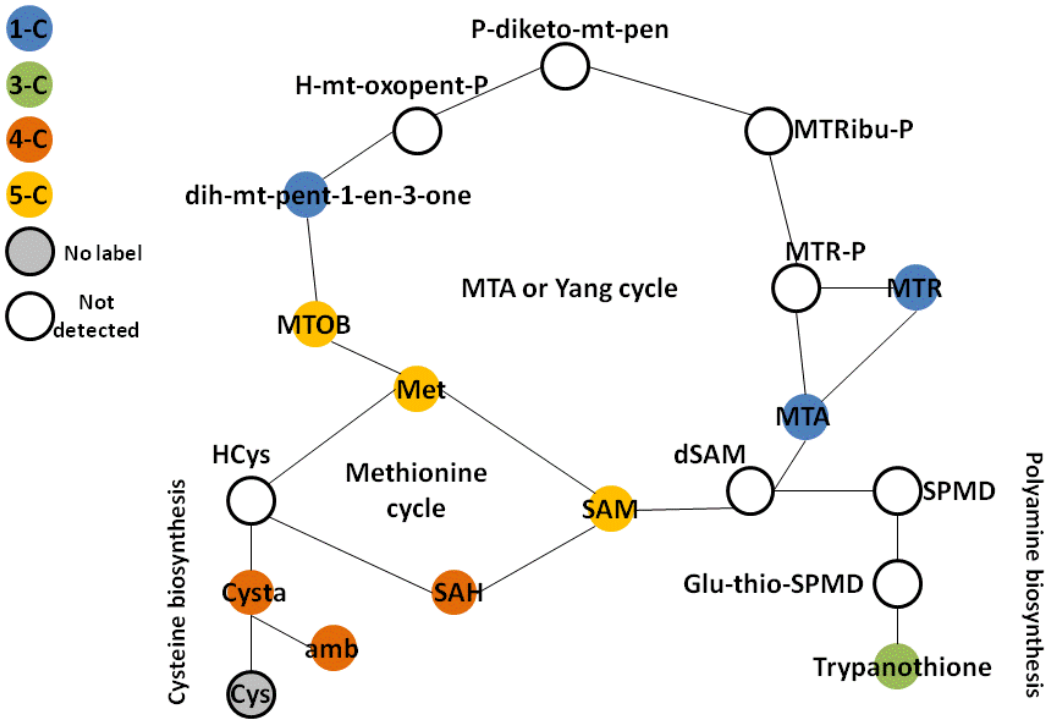


Figure 4.8: Overview of metabolites labelled from 13C L-methionine in bloodstream form trypanosomes. This basic metabolic network shows the metabolites mapped into their pathway (according to metacyc). Yellow nodes indicate 5 carbon labelling, orange nodes: 4 carbon labelling, green nodes: 3 carbon labelling, blue nodes one carbon labelling and none carbon labeeling is presented in grey nodes. Metabolites expected to be there but not detected are marked by white nodes. Met: L-methionine, SAM: S-adenosyl-L-methionine, SAH: S-adenosyl-L-homocysteine, HCys: L-homocysteine, Cysta: Cystathionine, amb: (2Z)-2aminobut-2-enoate, Cys: L-cysteine, dSAM: decarboxylated S-adenosyl-L-methionine, SPMD: Spermidine, Glu-thio-SPMD: glutathionyl-spermidine, MTA:5'-methylthioadenosine, MTR: 5'-methylthioribose, MTR-P: 5'-methylthioribose-1-phosphate, MTRRibu-P: 5'-methylthioribulose-1-phosphate, P-diketo-mt-pen: 1-Phospho-2,3-diketo-5'-methylthiopentane, H-mt-oxopent-P: 2-hydroxy-5-(methylthio)-3-oxopent-1-enyl phosphate, dih-mt-pent-1-en-3-one: 1,2-dihydroxy-5-(methylthio)pent-1-en-3-one, MTOB: 2-oxo-4-methylthiobutyrate.

4.2.1.1. Methionine cycle and L-cysteine biosynthesis

L-methionine was added to cultured BSF trypanosomes in a 50% mixture of ^{12}C and U-^{13}C-L-methionine for 48 h. The intracellular L-methionine content was approximately 35%. In fresh CMM this amount was about 30%. Although, the added L-methionine to the media was 50% labelled/unlabelled, the added FBS gold seems to dilute the L-methionine concentration in the fresh media. From the four metabolites that make the methionine cycle only three were detected in this dataset (Figure 4.9).

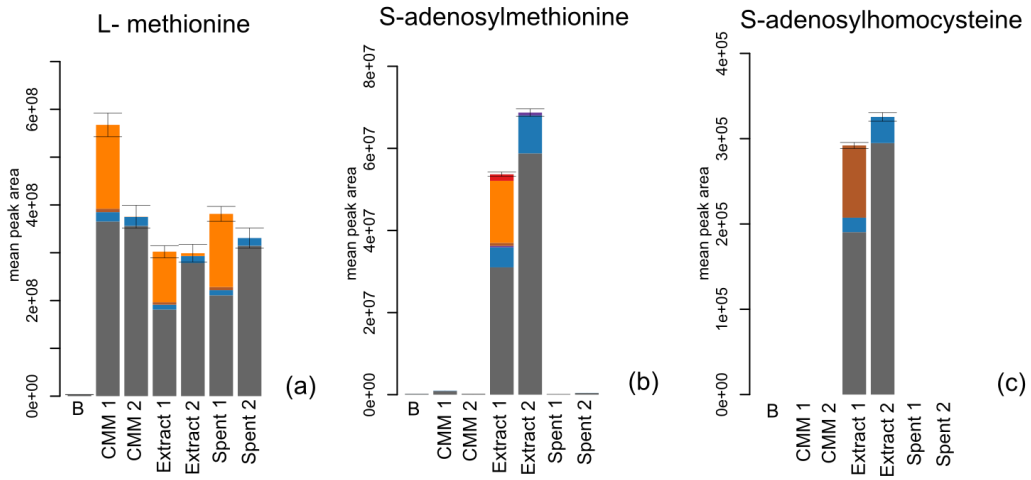


Figure 4.9: Labelling trend of the first three metabolites of the methionine cycle. (a) L-methionine, 5-C labelled in fresh media with 50% added U-^{13}C L-methionine, trypanosome cell extract and spent media; (b) S-adenosylmethionine, detected in cell extract, 5-C labelled in cell extract from 50% U- ^{13}C L-methionine culture; (c) S-adeosylhomocysteine, 4-C labelled in trypanosome extract incubated in CMM with 50% U-^{13}C L-methionine. CMM: Fresh medium sample, Extract: Trypanosome cell extract and Spent: Spent medium analysis. Numbers indicate samples with and without ^{13}C L-methionine (1 with ^{13}C compound, 2 without).Orange equals 5-C labelled, brown 4-C labelled and blue 1-C labelled carbons. As natural abundance of ^{13}C occurs, when 1-C is detected it needs to be compared to the unlabelled control.

As shown in Figure 4.10 formation of S-adenosylmethionine could be detected from L-methionine, with the labelling pattern as expected (5-C). S-adenosylmethionine to S-adenosylhomocysteine showed a resulting labelling pattern of 4-C for S-adenosylhomocysteine. Methylated compounds were also detected and are shown in 4.2.1.4. Conversion of S-adeosylhomocysteine to homocysteine would show a labelling of 4-C, however L-homocysteine was not detected (the authentic standard was also not detected). The formation of L-methionine from L-homocysteine could not be detected. Although 4-C labelled L-methionine was detected in the cellular metabolome, it was only shown in small amounts with a very similar percentage of 4-C labelled methionine in the fresh culture medium. Therefore, it seems that the 4-C labelled L-methionine originates from a small fraction of 4-^{13}C present in the supplied U-^{13}C L-methionine. Therefore it seems unlikely that the full methionine cycle is active in cultured bloodstream form trypanosomes or just with a very low flux, when large quantities of methionine are presented to the cell. Cysteine biosynthesis can occur from L-homocysteine, which has been produced from L-methionine via the methionine cycle. L-homocysteine and L-serine form L-cystathionine (catalysed by cystathionine β-synthase), which can be transformed to L-cysteine and (2Z)-2-aminobut-2-enoate by cystathionine γ-lyase. L-homocysteine would be expected to show a 4-C labelling pattern. L-cystathionine did show a 4-C labelling pattern expected when labelling occurs from L-homocysteine. As L-cysteine only gets the sulphur-group from L-cystathionine, I cannot say for certain if L-cysteine is being

produced from L-methionine. However, a peak was detected and putatively annotated as 2-aminobut-2-enoate which shows the expected labelling pattern from this reaction. It was shown by (Bacchi et al., 1995) that when ^{35}S-methionine was added to culture medium the ^{35}S-label shows up in L-cysteine, confirming that this pathway is active in *T. brucei*.

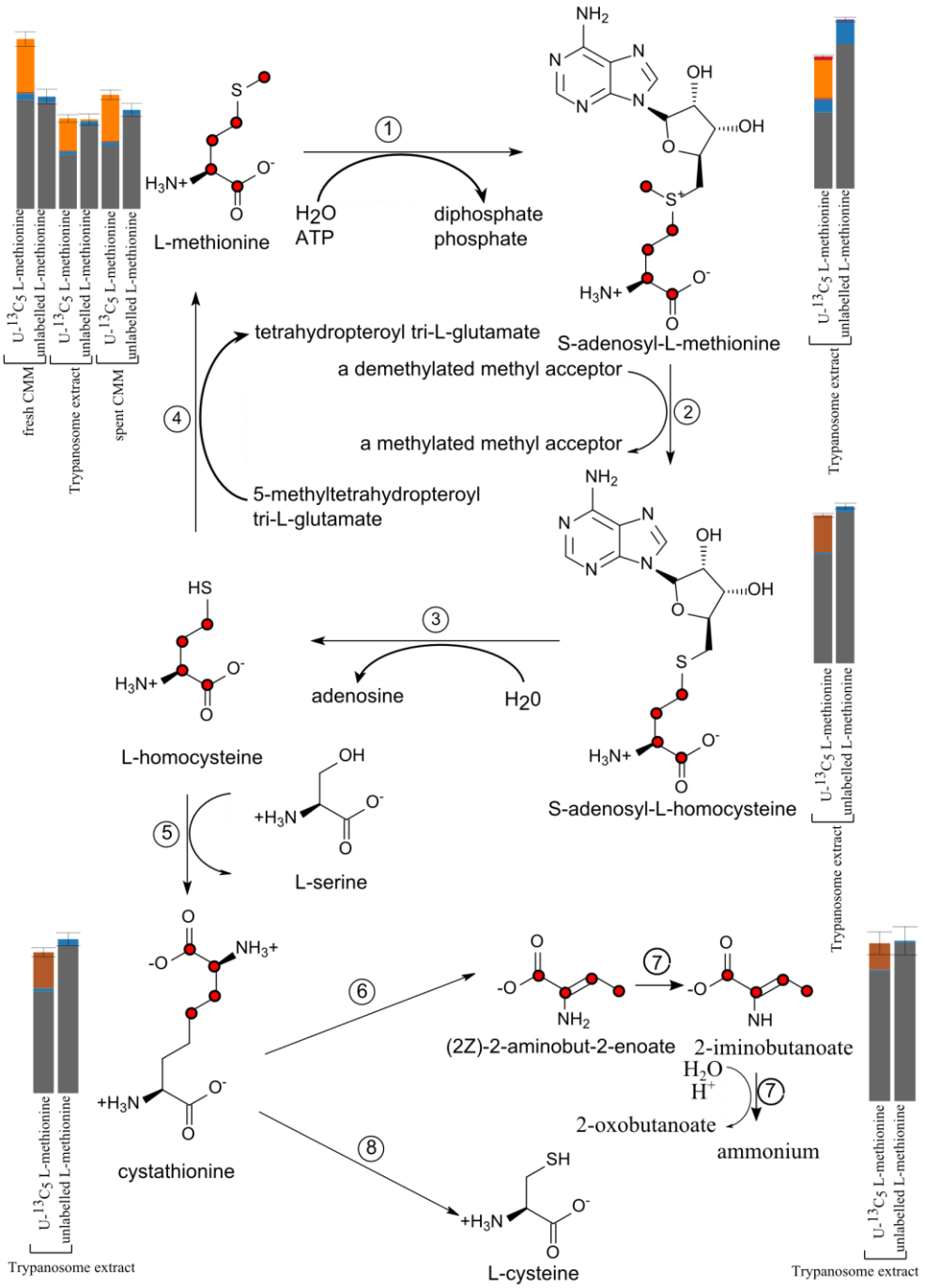


Figure 4.10: Methionine cycle and L-cysteine biosynthesis in trypanosomes as seen from labelling data. Expected labelled carbons are shown in red and when compound was detected in the dataset, labelling pattern is shown next to the structure. Orange equals 5-C labelled, brown 4-C labelled and blue 1-C labelled carbons. As natural abundance of ^{13}C occurs, when 1-C is detected it needs to be compared to the unlabelled control. Key to enzymes: 1- S-adenosylmethionine synthetase; 2-various methyltransferase; 3- S-adenosylhomocysteine hydrolase; 4- 5'-Methyltetra-hydrofolate:homocysteine methyltransferase (methionine synthase). 5-Cystathione β-synthase (EC 4.2.1.22), 6/8-γ-Cystathionase (EC 4.4.1.1), 7- spontaneous reaction.

4.2.1.2. Polyamine biosynthesis

The conversion of S-adenosylmethionine to decarboxylated S-adeosylmethionine (dSAM) is the first step in the synthesis of polyamines from L-methionine. dSAM was not detected in this dataset. dSAM and putrescine get converted to spermidine. However, spermidine does not retain well on ZIC-pHILIC columns. As seen with the authentic standard of spermidine (see Figure 4.11) no defined peak is detectable, but many over the time period of about 7 to 20 min. Therefore, no identification of the compound spermidine can be made for analysis.

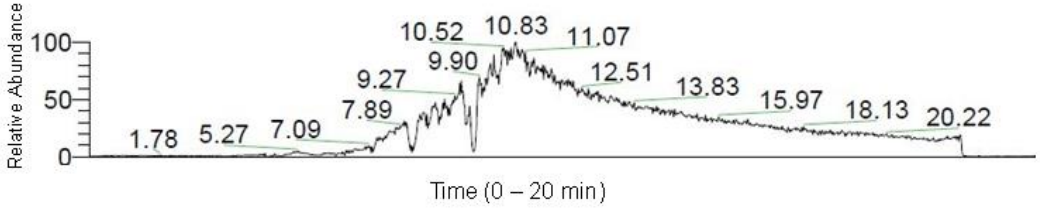


Figure 4.11: Spermidine from authentic standard mix on ZIC-pHILIC column.

The only metabolite detected in this pathway was trypanothione disulfide, which was 3-C labelled from methionine. As seen in Figure 4.12 , the detected labelling pattern for trypanothione is shown to be 1-C, 3-C and 4-C labelled. The detected 4-C labelled is most likely an artefact from the 3-C originating from L-methionine and the natural occurring 1-^{13}C.

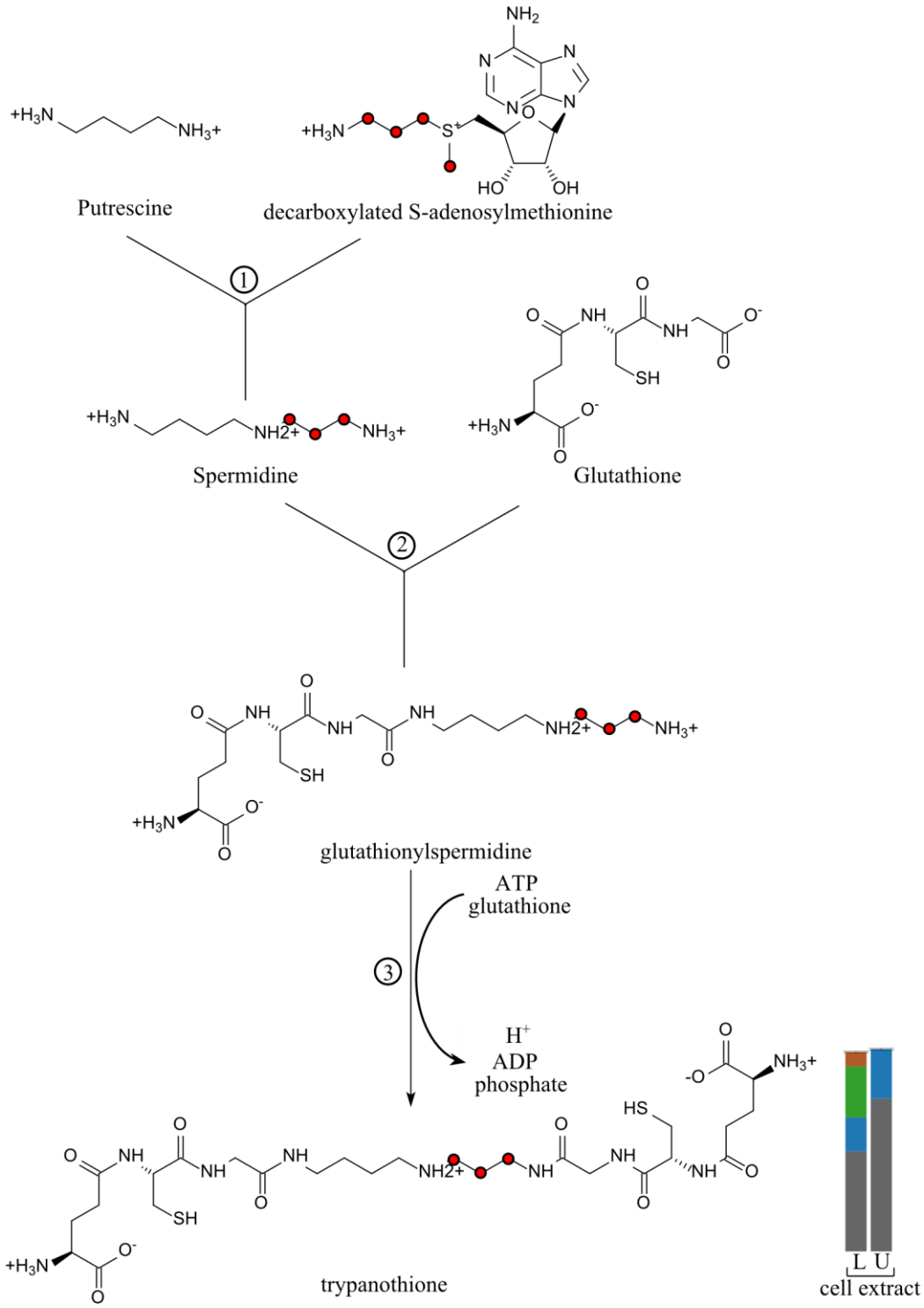


Figure 4.12: Polyamine biosynthesis from L-methionine. dSAM and putrescine are converted to spermidine and S-methyl-5'-adenosine by 1- Spermidine synthase (EC 2.5.1.16). Spermidine is utilised further in the synthesis of polyamines, while S-methyl-5'-adenosine enters the L-methionine salvage pathway (Yang cycle). Formation of glutathionylspermidine from spermidine and glutathione is catalysed by 2- Glutathionylspermidine synthetase (EC 6.3.1.8). Glutathionylspermidine and glutathione form trypanothione, catalysed by 3- Trypanothione synthase (EC 6.3.1.9). Expected labelled carbons are shown in red and when compound was detected in the dataset, labelling pattern is shown next to the structure. Brown equals 4-C labelled, green 3-C labelled and blue 1-C labelled carbons.

4.2.1.3. Recycling of methionine

It has been previously described in trypanosomes that methionine can be salvaged via the methylthioadenosine (MTA) pathway (Berger et al., 1996; Nozaki et al., 2005). S-methyl-5-thioadenosine (1-C labelled from L-methionine) being converted back to L-methionine, going through a seven step pathway including 5'-methylthioribose-phosphate, 5'-methylthioribulose-1-phosphate, 1-phospho-2.3-diketo 5'-methylthiopentane and 2-oxo-4-methylthiobutyrate (see Figure 4.14.). However, in this labelling data only the intermediates S-methyl-5-thioadenosine and methylthioribose are being labelled. L-methionine was detected in 1-C labelled form. However, that seems to be attributed to the normal natural abundance detection. Figure 4.9 (a) shows the detected L-methionine plus labelling pattern and there is no increase in 1-C labelled L-methionine between cells incubated in media with and without 13-C labelled L-methionine.

Analysis of spent media from bloodstream form trypanosomes have shown, that methylthioribose seems to be secreted from the cells after 48 hours (D. H. Kim and F. Achcar, unpublished data). Methylthioribose was also detected in spent media during this experiment, however, only one replicate of spent media was analysed and the peak was ill-defined. Nonetheless, together with the spent media analysis (Figure 4.13) data from Kim and Achcar it seems plausible that 5'-methylthioribose gets secreted from bloodstream form trypanosomes when L-methionine is highly abundant in the culture medium.

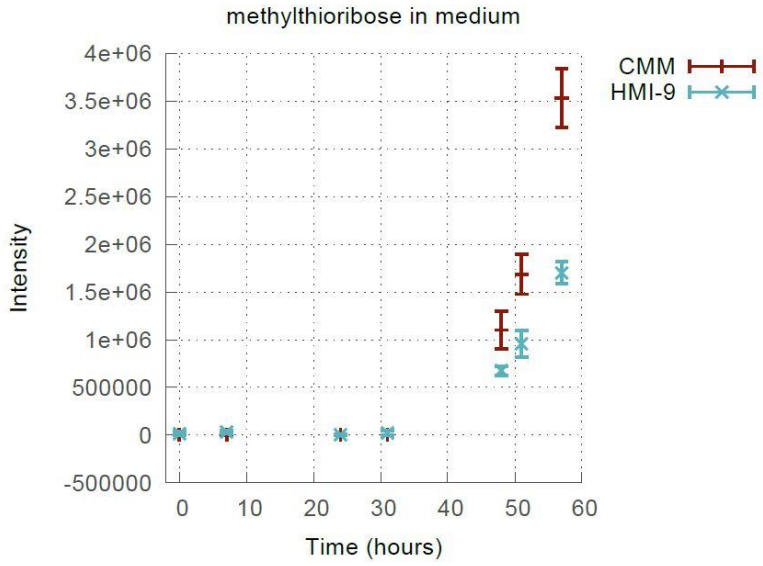


Figure 4.13: Spent media analysis (performed by D. H. Kim and F. Achcar), shows increasing levels of methylthioribose secreted from the cells, starting after 48h. Trypanosomes were grown in HMI-9 and CMM. Figure kindly provided by F. Achcar

Creek et al. (2015) showed that when bloodstream form trypanosomes were incubated with ^{13}C-labelled Glucose, S-adenosylmethionine got labelled from ATP and those labelled carbons were transferred to MTA. However, from that point onwards no labelled carbons were detected.

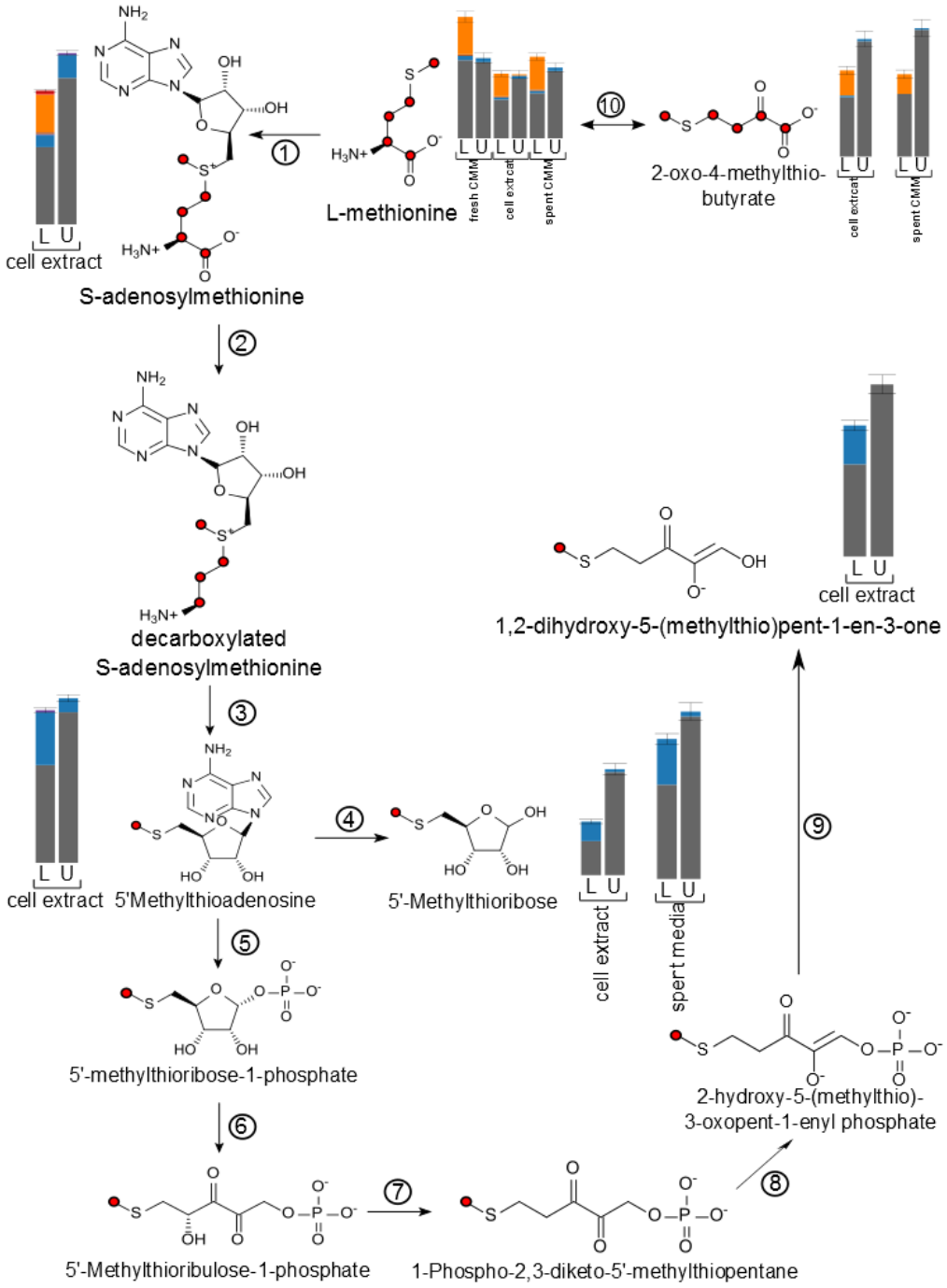


Figure 4.14: Methionine salvage pathway via 5'Methylthioadenosine and 5'Methylthioribose. Enzyme key: **1**-S-adenosylmethionine synthetase, **2**-SAM decarboxylase, **3**-spermidine synthase, **4**-methylthioadenosine nucleosidase, **5**-methylthioadenosine phosphorylase, **6**- 5-methylthioribose-1-phosphate isomerase, **7**- 5-methylthioribulose-1-phosphate dehydratase, **8**- 2,3-diketo-5-methylthio-1-phosphopentane enolase, **9**-2-hydroxy-3-keto-5-methylthio-phosphopentene phosphatise, **10**- transaminase. Expected labelled carbons are shown in red and when compound was detected in the dataset, labelling pattern is shown next to the structure. Red equals 6-C labelled, orange 5-C labelled and blue 1-C labelled carbons.

Berger et al (1996) reported that the last step in methionine production via methylthio-adenosine takes place, converting 2-oxo-4-methylthiobutyrate into L-methionine. However, this is inconsistent with the conclusion of this experiment. A compound putatively annotated as 4-methylthio-2-oxobutanoate with 5-C labelling was detected in the ^{13}C l-methionine dataset in cell extract and spent media samples. So it seems more likely that 2-oxo-4-methylthiobutyrate is being produced from L-methionine and secreted from the cell, rather than the MTA cycle being active and salvaging L-methionine from methylthioadenosine. However, 5'methylthioadenosine phosphorylase activity was measured in *T.b. brucei* crude extract (Ghoda et al., 1988) highlighting the presence of the methylthioadenosine pathway. A compound putatively identified as 1,2-dihydroxy-5-(methylthio) pent-1-en-3-one was found 1-C labelled. This compound is described as an intermediate of the Yang cycle for L-methionine salvage and can either be a precursor to 4-methylthio-2-oxobutanoate or 3-(methylthio)propanoate (Figure 4.15 (b)). This reaction has been described to happen in bacteria, and is thought to provide a mechanism for L-methionine regulation *in vivo* (Dai et al., 2001).

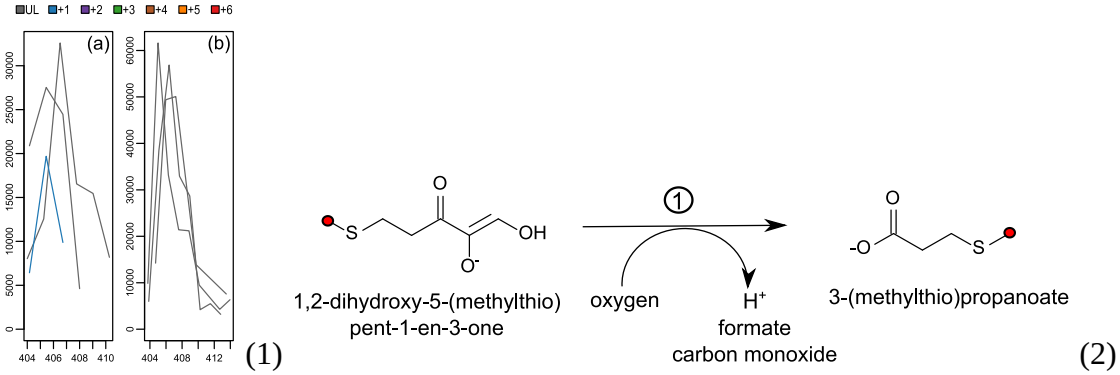


Figure 4.15: 1,2-Dihydroxy-5-(methylthio) pent-1-en-3-one (1) Detected peak for 1,2-dihydroxy-5-(methylthio) pent-1-en-3-one. Although the peak is ill-defined, the 1-C labelling appears clear in cell extract grown in 50% U-^{13}C L-methionine for 48 hours (a) and not in the cell extract lacking the U-^{13}C L-methionine (b). (2) Possible fate of 1,2-dihydroxy-5-(methylthio) pent-1-en-3-one in bloodstreamform trypanosomes. 3-(methylthio)-propanoate was not detected. enzyme key: 1- acireductone dioxygenase.

4.2.1.4. Methylated metabolites

Six methylated compounds with ^{13}C labelling were detected in this dataset. Labelled methylated amino acids with 1-C, 2-C and 3-C labelled and 1-C labelled methylguanine.

Various methyltransferases use S-adenosylmethionine as a methyl donor and their targets include tRNA, proteins, DNA and lipids. However, those compounds are too big to be analysed on this platform. 41 genes are annotated as putative methyltransferases in the published *T. brucei* genome and about 90% of the SAM produced in trypanosomes is thought to be used for methylation processes. However, as for this study the methylated products were seen as by-products of the metabolic pathways associated with L-methionine, it was not further investigated. The six compounds were:
N6,N6, dimethyl-L-arginine, N6 methyl L-arginine, N6 methyl L-lysine, N6, N6, N6 trimethyl L-lysine, N-methylhistidine and 7-methylguanine (see Figure 4.16)

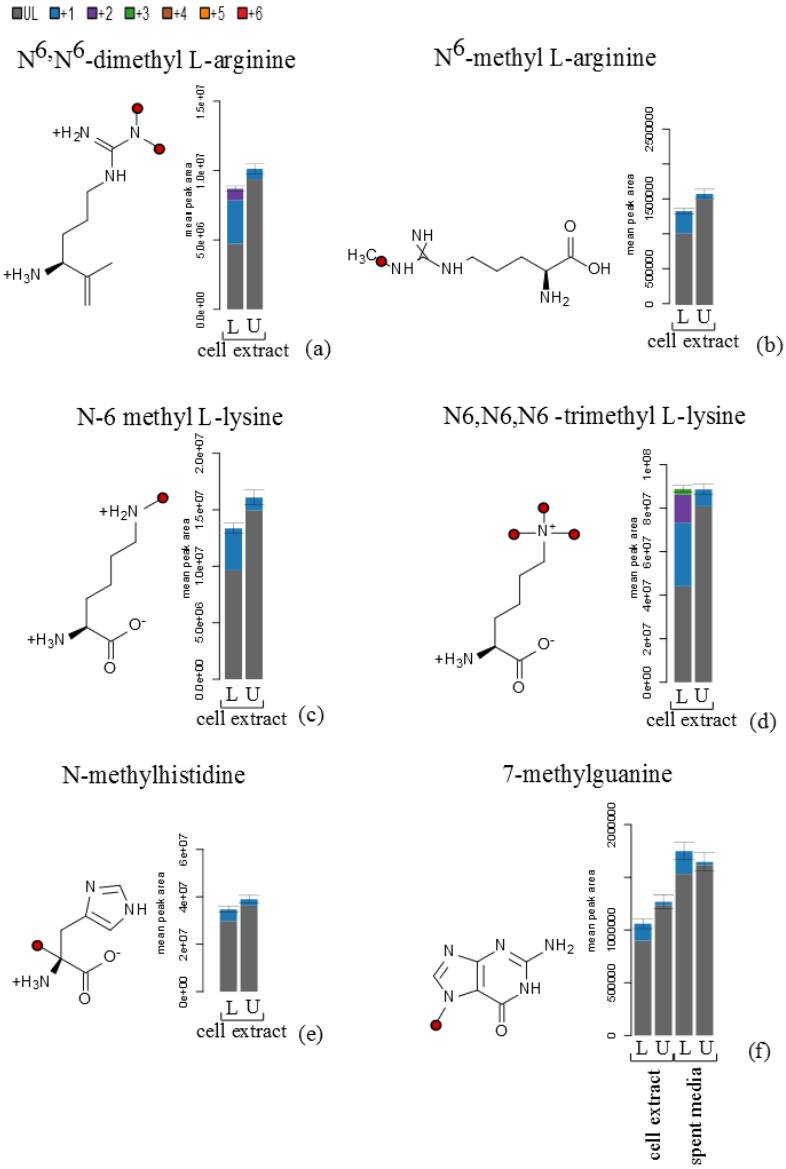


Figure 4.16: Methylated metabolites detected in trypanosome cellextract, incubated with (L) or without (U) ^{13}-C L-methionine.

4.2.2. U- ^{13}C - L-arginine metabolite tracking

Trypanosomes can take up L-ornithine and L-arginine, but potentially the L-ornithine uptake could be sufficient enough for trypanothione synthesis (I. Vincent, PhD thesis, 2011), which could explain the absence of an arginase in trypanosomes. However, as it was shown that L-arginine can be converted to L-ornithine, when L-ornithine was absent from the incubation medium (I Vincent, thesis) an alternative route for L-ornithine production from L-arginine could be present. To further probe the L-arginine metabolism in trypanosomes, U-^{13}C-L-arginine was used in an untargeted metabolomic approach. 13 labelled compounds were detected in U-^{13}C labelled L-arginine dataset, six of which were related to the metabolites shown and mapped in Figure 4.7.

4.2.2.1. Urea cycle – Biosynthesis of L-ornithine

L-arginine was found 6-C labelled in fresh CMM (with 200µM U-^{13}C L-arginine added), labelled cell extract, spent labelled CMM, but also in small amounts in unlabelled cell extracts. The most reasonable explanation for finding labelled L-arginine in unlabelled cell extract is sample contamination, either during sample preparation or sample carry-over on the LC-MS platform. After checking the sample running order on the LC-MS platform it was shown that unlabelled cell extract samples (three replicates) were run directly after samples containing labelled L-arginine. As the L-arginine peaks are the only peaks showing contamination it is likely that the relatively high concentration of L-arginine caused minimal contamination to show up in the datsaset, while possible other contaminations were too low to show. L-ornithine was detected 5-C labelled from L-arginine. It is unclear whether the reaction happened within the cells or in the medium. Spent media showed equally high amounts of L-ornithine 5-C labelled as cell extract, while fresh media only showed traces of 5-C labelled L-ornithine. However, since other metabolites labelled inside the cell do not show outside and the concentration of labelled L-ornithine was similar in cell extracts and in spent medium, it seems L-arginine converts to L-ornithine outside the cell by an arginase found in the added serum. A low percentage of 6-C labelled L-citrulline was detected in the labelled cell extract. This labelling pattern could only occur directly from L-arginine, as L-ornithine was 5-C labelled. The intermediate metabolite L-arginino-succinate was not detected in this dataset. N-acetyl-L-ornithine and N-acetyl-L-glutamate 5-semialdehyde were detected as 5-C labelled from L-arginine. They are pathway intermediates in the biosynthesis of L-ornithine from L-

glutamate. L-glutamate was not labelled in this dataset so the pathway does not occur to be fully functional. The enzyme Acetylornithine deacetylase, which converts L-ornithine to N-acetyl-L-ornithine, has been putatively identified in the trypanosome genome. The labelling pattern for N-acetyl-L-glutamate 5-semialdehyde was ill-defined (see Figure 4.17.). Furthermore, the enzyme Acetylornithine transaminase has not been reported to be present in *T. brucei*, so it is unclear if the detected labelling is indeed from N-acetyl-L-glutamate 5-semialdehyde. Also, traces of labelled N-acetyl-L-glutamate 5-semialdehyde were detected in the cell extract from cultures without added ^{13}C L-arginine.

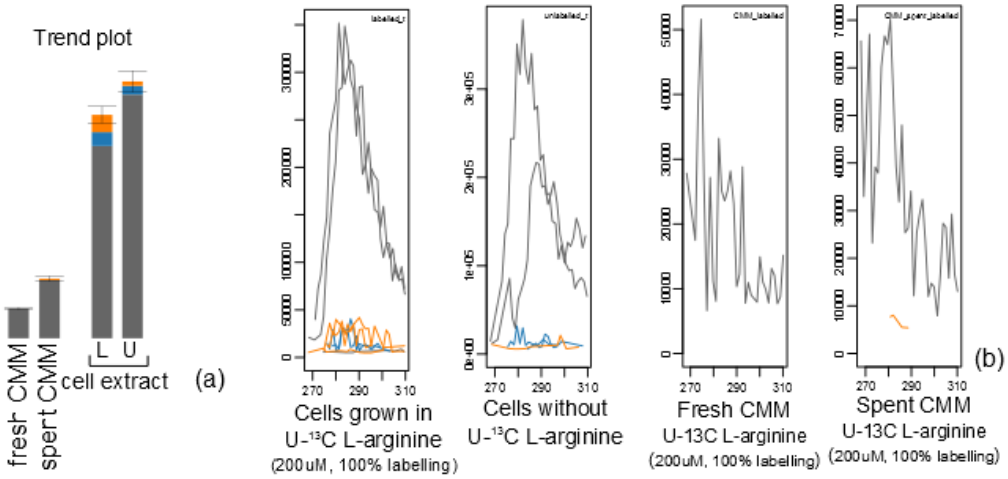


Figure 4.17: N-acetyl-L-glutamate 5-semialdehyde (a) Trend plot of 5-C labelled N-acetyl-L-glutamate 5-semialdehyde in fresh and spent media (with added U-^{13}C L-arginine) and labelled and unlabelled cell extract. Orange equals 5-C labelled and blue 1-C labelled carbons. As natural abundance of ^{13}C occurs, when 1-C is detected it needs to be compared to the unlabelled control. (b) Detected peaks from those samples.

From L-ornithine, the production of trypanothione (4-C labelled) could be detected, although the intermediates putrescine, spermidine and glutathionylspermidine were not. However, putrescine, spermidine and glutathionylspermidine do not retain well on zic-pHILIC columns. All the metabolites described here with their expected and actual labelling pattern are shown in Figure 4.18.

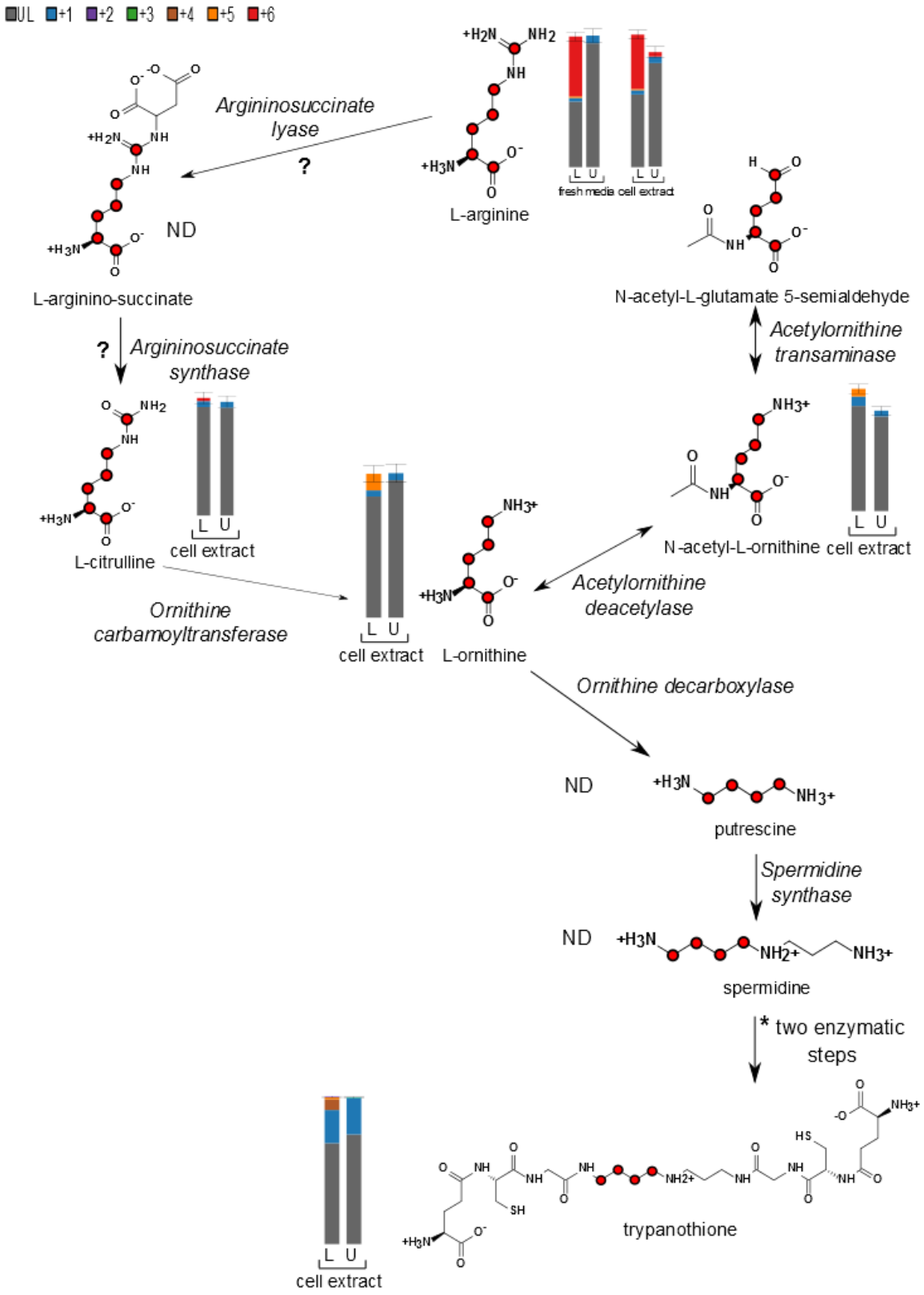


Figure 4.18: Predicted metabolites to be involved in L-ornithine biosynthesis and degradation (Urea cycle and polyamine biosynthesis included). Carbons labelled from L-arginine are highlighted in red and enzyme names are in italic. * Two enzymatic steps are shown in Figure 4.4 and 4.5.

4.2.2.2. Energy storage in T. brucei

L-arginine phosphate (or phosphoarginine) is important for trypanosomes for energy storage. In this experiment phosphoarginine was found to be labelled from L-arginine at 38% (see Figure 4.19), when about 45% labelled L-arginine was detected in fresh media and 40% labelled L-arginine within the cell.

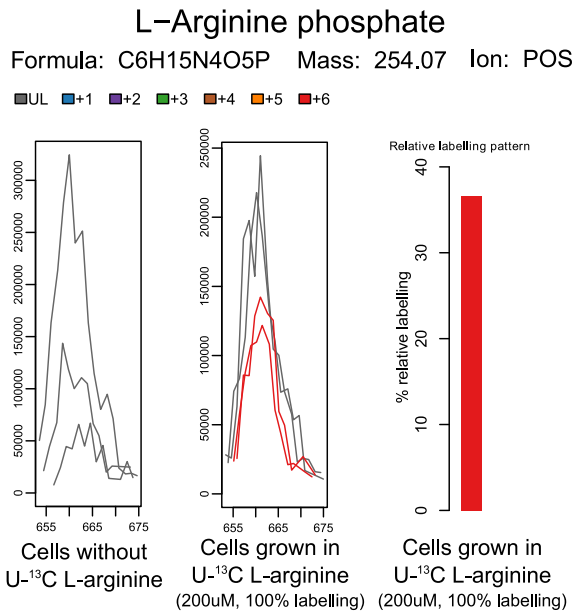


Figure 4.19: Detected peak of L-arginine phosphate, with a mass of 254.07 and retention time of: 11.08 min.

4.2.2.3. Other metabolites labelled from L-arginine

- Methylarginine and dimethylarginine, both 6-C labelled from L-arginine, could have their origin from methylated proteins.

  A recent study highlights the importance of arginine methylated proteins. Five putative arginine methyltransferases in *T. brucei* genome have been described and proteomic analysis showed 167 arginine methylproteins which have a wide range of function, including metabolism, chaperoning, RNA processing, DNA replication, translation and function in transport. Some of those methylations are trypanosome specific, while others are conserved modifications (Fisk et al., 2013)
- Ketoarginine, which has also been identified as 6-C labelled from L-arginine, could be a precursor for 3-methylarginine synthesis or L-arginine degradation.
- N5-(L1-Carboxyethyl)-L-ornithine formed from L-ornithine and pyruvate. This compound was seen to be increased after cells were treated with Eflornithine

(Vincent et al., 2012). Additionally, N5-(L1-Carboxyethyl)-L-ornithine was also detected in U-^{13}C labelled Glucose data (Creek et al.,2015). Although this is a metabolite known from bacteria, its function in trypanosomes is unknown (reaction, see Figure 4.20 (a)).

- A metabolite with the mass of 173.069 and a retention time of 9.7 min, putatively identified as L-N2-(Carboxyethyl)-L-arginine (D-octopine). D-octopine plays a role in NAD/NADH regulation and is formed by D-octopine synthase using pyruvate and L-arginine (Figure 4.20 (b)).

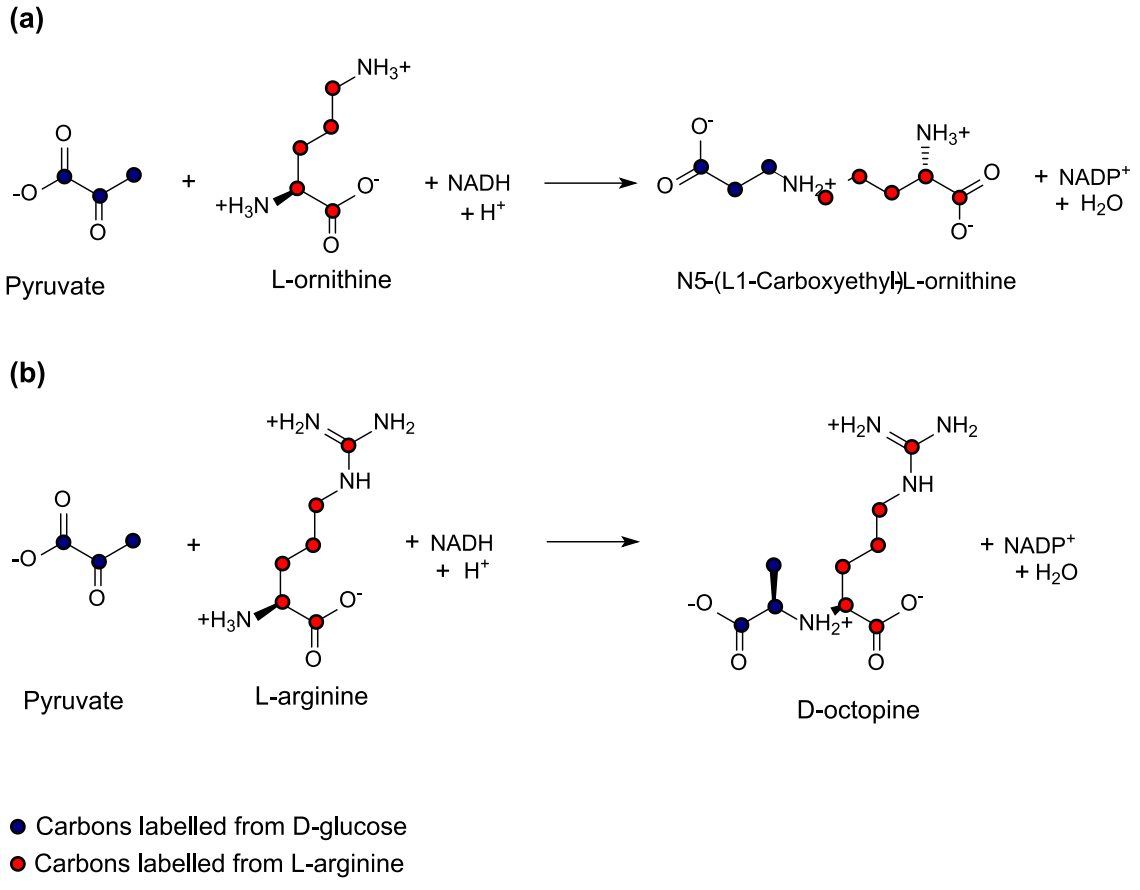


Figure 4.20: Labelling pattern and origin of carbons in reaction leading to (a) N5-(L1-Carboxyethyl)-L-ornithine and (b) D-ocyopine. (a) Enzyme catalysing the reaction from pyruvate and L-ornithine to N5-(L-1-Carboxyethyl-) L-ornithine is N5-(carboxyethyl)ornithine synthase (EC1.5.1.24). (b) Synthesis of D-octopine by octopine dehydrogenase (EC1.5.1.11) from pyruvate nd L-arginine. Blue dots indicate carbon originated from D-glucose (Creek et al, 2015) and red dots indicate carbons labelled from L-arginine.

4.2.3. U- ^{13}C - L-proline metabolite tracking

Although L-proline seems to be taken up by bloodstream form trypanosomes, the intracellular concentration of labelled L-proline was very low using 50% of labelled L-proline (50 µM final concentration U-^{13}C L-proline) in growth medium for 48 hours. Under those conditions downstream metabolites from L-proline were not labelled to a high enough percentage for analysis and therefore the experiment was repeated with 100% added labelled L-proline (200 µM U-^{13}C L-proline). Recycling of L-proline from protein/peptide breakdown seems a plausible explanation for the dilution of labelled L-proline within the cell and would also explain why the first experiment was not successful.

Seven labelled compounds were detected in U-^{13}C labelled L-proline dataset, all with relatively low percentage of labelling (except for L-proline). Figure 4.7 shows the detected meatbolites including the percentage of labelling.

The results of this experiment conclude that except for protein synthesis, L-proline seems to be playing a minor role in the metabolism of bloodstream form trypanosomes. Trypanothione was detected to be 5-C labelled from L-proline, but only 1.2%. The pathway for this is shown in Figure 4.21 and the intermediates are (S)-1-pyrroline-5-carboxylate and L-glutamate-5-semialdehyde (not detected), L-glutamate (0.5% labelled,

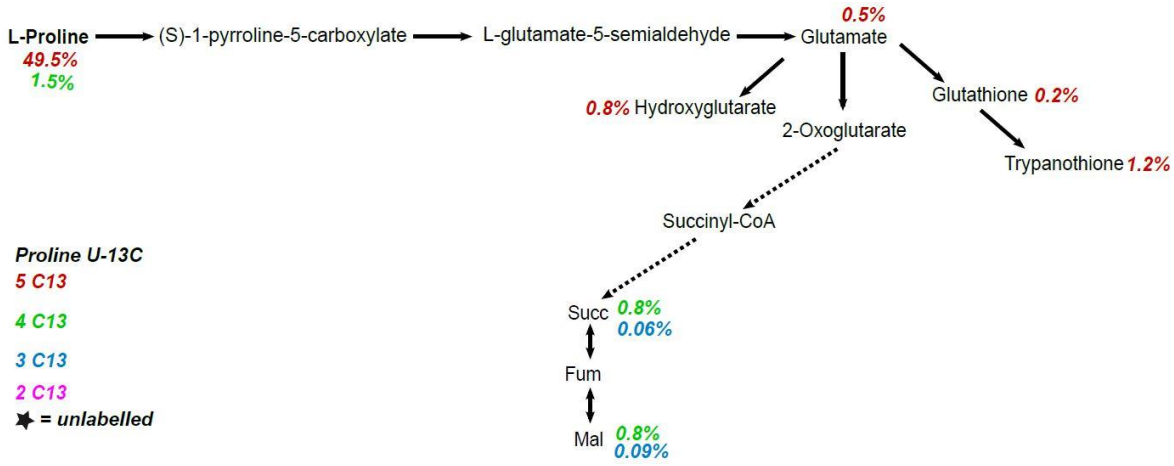


Figure 4.21: Distribution of labelled carbons (C13) from L-proline (Fiona Achcar).
Colours representing the number of carbons contributed from L-proline. Red: 5-C, green: 4-C, blue: 3-C and pink: 2C. Not detected compounds are (S)-1-pyrroline-5-carboxylate, L-glutamate-5-semialdehyde, 2-oxoglutarate and Succinyl-CoA.

5-C) and Glutathione (0.2% labelled, 5-C). The detected labelling of trypanothione therefore results from L-glutamate and glutathione, rather than L-ornithine and putrescine. Other metabolites labelled were hydroxyglutarate (0.8%) and succinate and malate (0.8% 4-C labelled both and 0.06% (succinate) and 0.09% (malate) 3-C labelled).

4.3. Discussion

4.3.1. L-methionine

Here, the sulphur containing amino acid L-methionine and its metabolism in *T. brucei* was investigated, with the focus on the methionine cycle, polyamine biosynthesis and the MTA cycle.

The methionine cycle shows three out of the four metabolites present. Although L-homocysteine was not detected, due to downstream metabolites being present (cystathionine) it only seems plausible that it would be 4-C labelled. Interestingly, the methionine cycle does not seem to be a cycle at all as the lack of 4C-labelled methionine indicates that L-homocysteine is not being converted back to L-methionine. This contradicts previous published material that L-homocysteine can be converted back to L-methionine (Walker and Barrett, 1997). It seems more likely that L-homocysteine gets converted to L-cysteine, over cystathionine (which showed 4-C labelling), in agreement with Bacchi et al (1995), who could show, with ^{35}S-labelled L-methionine, the reverse transulfation pathway being present in trypanosomes. The trypanosome specific polyamine thiol conjugate trypanothione was detected as being 3-C labelled from L-methionine. Unfortunately, no intermediate metabolite in this pathway was detected (dSAM, spermidine, glutathionylspermidine). However, methylthioadenosine (a byproduct of spermidine synthesis from dSAM and putrescine and starting point of the MTA cycle) was detected 1-C labelled. The production of methylthioribose was also detected in bloodstream form trypanosomes when incubated with ^{13}C glucose (Creek et al, 2015) and here it was 5-C labelled from S-adenosylomethionine (via ribose and the resulted labelling of ATP), which confirms the finding in the methionine labelled dataset.However, the full functionality of the MAT cycle could not be confirmed in either dataset. The production of methylthioribose from methylathioadenosine was detected, with both compounds 1-C labelled from L-methionine. But, it looks like methylthioribose gets secreted from the cell and not converted to methylthioribose 1-phosphate. Although the intermediates in the MTA pathway, from 5'methylthioribose-5-phosphate to 2-hydropxy-5-(methylthio)-3-oxopent-1-enyl phosphate, were not detected, the second last metabolite (1,2-dihydroxy-5-(methylthio)pent-1en-3-one) was. The 1-C labelling suggests that the pathway is active, but after that point is unclear what happens. 2-oxo-4-methylthiobutanoate is not 1-C labelled as would be expected. Instead 2-oxo-4-methylthiobutanoate is 5-C labelled from

L-methionine and seems to be secreted from the cell. As it is the corresponding keto-acid and it has been shown that trypanosomes are capable of doing this reaction (Creek et al., 2013). These results conflict with research from Berger et al (1996) and Bacchi et al (1991), who showed that *T. brucei* is capable of producing L-methionine from 2-oxo-4-methylthiobutanoate, preferably from the transamination of aromatic amino acids (Berger et al., 1996; Berger et al., 2001). However, as shown in the presented dataset, under the conditions used, L-methionine does not seem to be recycled by either methionine cycle or MTA cycle, which is consistent with findings from U-^{13}C glucose labelling (Creek et al., 2015).

4.3.2. L-arginine

L-ornithine is an important precursor for the synthesis of the diamine putrescine in *T. brucei*. Putrescine is involved in the synthesis of spermidine, which can be synthesised, with glutathione, to the trypanosome specific thiol trypanothione (Fairlamb and Cerami, 1992). Vincent et al (2011) also found the metabolite N-acetyl-L-ornithine being increased after bloodstream form trypanosomes were treated with Eflornithine. N-acetyl-L-ornithine is the precursor of L-ornithine in the biosynthesis of L-ornithine from L-glutamate. Tracking metabolites using stable isotope labelling has been successfully applied before, most recently with 13C-labelled glucose in bloodstream form trypanosomes, resulting in a much better understanding of the glucose metabolism (Creek et al., 2015). It has been reported that bloodstream form trypanosomes can convert L-arginine into L-ornithine, however without the production of urea (Vincent, thesis). Using ^{13}C L-arginine for metabolite tracking could lead to the identification of a different pathway for L-ornithine biosynthesis is active in bloodstream *T. brucei* that excludes the most direct route via 'arginase'.

L-ornithine was detected as 5-C labelled from Arginine in this experiment (see 4.2.2.1). However, it seems that this conversion happens in the media. Possible bovine arginase activity in the media could explain the 5-C-labelled L-ornithine. Labelled L-ornithine was detected in spent media in similar high concentration (about 10%) as within the cell (about 12%). Fresh media with added U-^{13}C L-arginine only showed traces of 5-C Ornithine, however, this sample was prepared immediately after production and therefore the bovine arginase activity was stopped very quickly. As bloodstream form trypanosomes can take L-ornithine up this seems the most reasonable explanation for the presence of 5-C-Ornithine within the cell.

Incubating U-^{13}C L-arginine in cell free media with 10% FBS also showed 5-C labelled L-ornithine after 48 hours (F. Giordani, data not shown), concluding it is most likely that L-arginine gets converted to L-ornithine in the media by arginase activity from serum. However, L-citrulline has been detected as 6-C labelled, albeit at small levels. L-citrulline is part of the urea cycle, a precursor in L-arginine production from L-ornithine. It is possible that in bloodstream form trypanosomes the urea cycle is operative in the reverse direction; leading from L-arginine via L-citrulline to L-ornithine and this data could support this hypothesis. L-citrulline was 6-C labelled, which could only occur from L-arginine, not L-ornithine. The intermediate L-arginino-succinate was not detected in the dataset.

L-arginine phosphate has been described to be of use in energy storage in trypanosomes, with three arginine kinases being described in the *T. brucei* genome.
The main phosphagen for energy storage in vertebrates is creatine, but trypanosomes do not seem to encode a creatine kinase and use arginine phosphate instead (Canepa et al., 2011).

4.3.3. L-proline

The percentage and labelling pattern from L-proline suggests that L-proline is only minimally involved in the synthesis of polyamines. The data shows that L-proline is not involved in the synthesis of L-ornithine, but rather transfers carbons via L-glutamate and glutathione to trypanothione. However, the low percentage of labelled compounds detected in labelled cells, imply that the main purpose for L-proline in bloodstream form trypanosomes is for protein synthesis.

Chapter 5

5.1. Introduction

The two described causative subspecies of HAT are *T. b. rhodesiense* (acute form) and *T. b. gambiense* (chronic form), with both forms present in the Central Nervous System (CNS) in the late stage of the disease. For *T. b. brucei* infections in mice, it has been shown that different strains of trypanosomes can manifest themselves as acute and chronic diseases as well. In this project, the *T. b. brucei* strains GVR35 (Jennings et al., 2002) and 427 (Melville et al., 2000) were used, as GVR35 causes a chronic infection in mice and invades the CNS with varying parasitemia while 427 shows an acute infection with high parasitemia causing high mortality within days without invading the CNS (see 1.2.3). What causes this difference in the progression of infection? Secreted/excreted proteins from the parasites (secretome) have been described as being important factor for virulence and to avoid the host immune response (Geiger et al., 2010). Garzon *et al.* showed that excreted/secreted proteins can inhibit the maturation of dentritic cells and stop them from inducing a lymphocytic allogenic response (Garzón et al., 2013). Furthermore, secreted cysteine peptidases are thought to be an important factor in the crossing of the blood brain barrier (BBB). Analysing and comparing the secretome from both strains may therefore give an inside of what causes the difference in virulence between 427 and GVR35.

Two approaches were used to determine the secretome of the two different *T. b. brucei* strains: (1) differential secreted proteins between the two strains were investigated by DiGE and (2) the whole set of secreted proteins for each strain using Filter Aided Sample Preparation (FASP) with in solution trypsin digest.

5.2. Results

Two strains of bloodstream form trypanosomes were grown up from stabilates. To keep the conditions for both strains the same, different media (HMI-9, CMM and modified HMI-9, see 2.1.1.) were tested for cell culture. Strain GVR35 was directly grown in modified HMI-9 medium, while strain 427 was passaged once in HMI-9 before changed to modified medium or both strains were adapted to CMM + 10% FBS gold. However, strain GVR35 did not grow in HMI-9 + 10% FBS gold. Adaption to the modified media took several

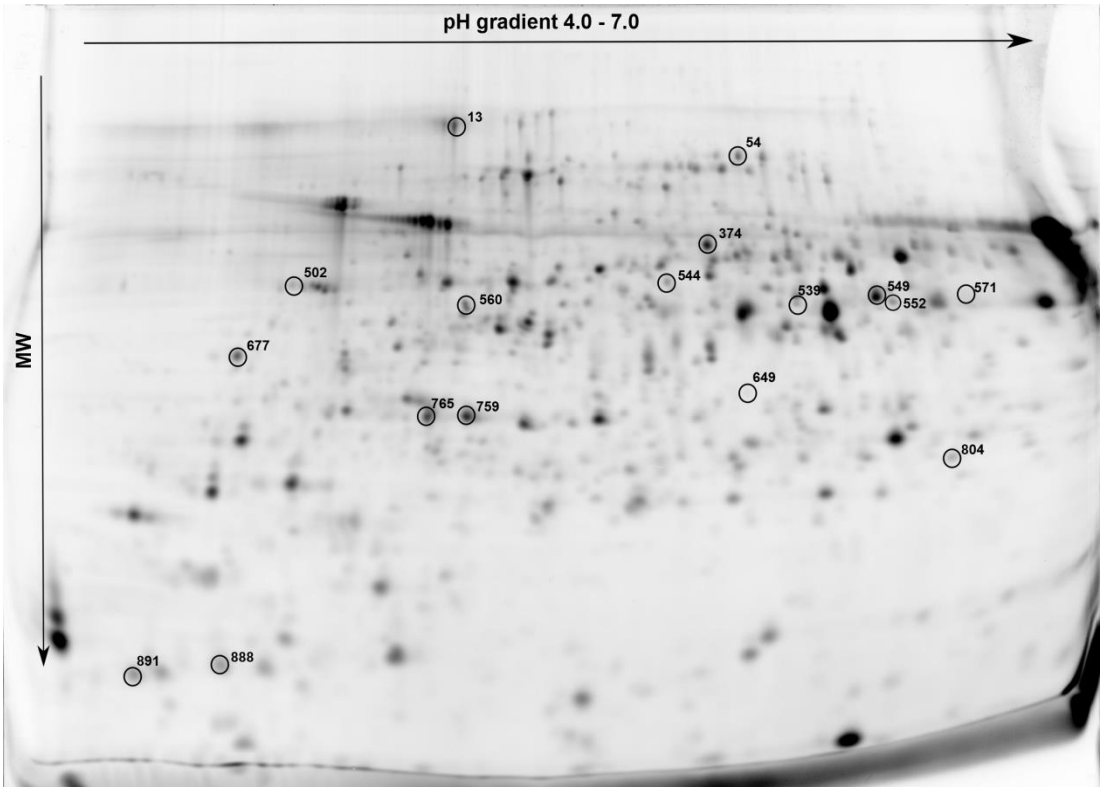
passages for strain 427 or strain GVR35 to adapt to CMM. To reach the desired cell count of $2x10^8$ (as described by Holzmueller *et al.*, 2008) cell culture volumes were between 120 - 150 ml for strain 427 and 150 - 170 ml for strain GVR35. Cell viability was checked using microscopic examination during incubation in serum-free modified medium every 30 min for two hours. Cell morphology looked normal in strain 427 and cells had good motility. Strain GVR35 showed cell clumps forming after 60 min, which suggests that cells were starting to die. After secretome sample preparation an aliquot of the sample was run on SDS-PAGE to ensure a detectable amount of protein was present.

5.2.1. Sample preparation for DiGE

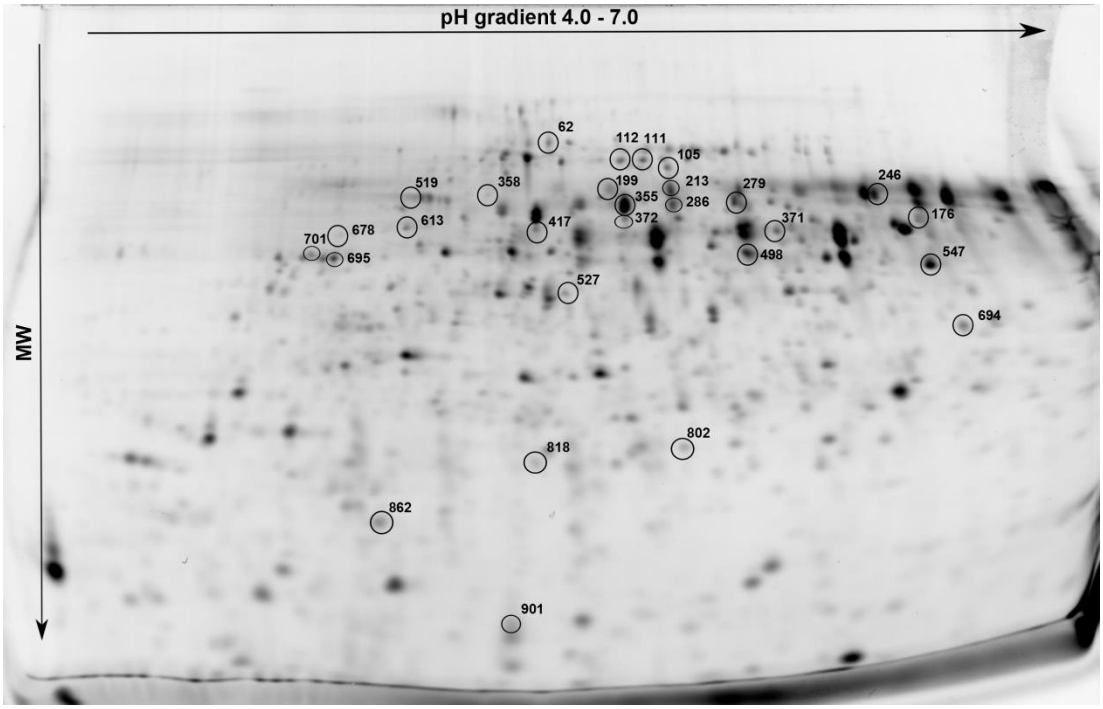
For quantification purposes and to determine the strain specific differences in secreted proteins a gel-based proteomic approach was tested. 2D-DiGE is an excellent tool to compare different strains of pathogens. Samples are run on one gel, which minimises the chances of seeing dissimilarities in proteins detected due to differences in gel running conditions. A drawback of this method is that a large amount of protein is needed to run the 2D-DiGE with two prep-gels. 550 µg protein is required per sample set to obtain optimal results. The amount of protein obtained from the secretome preparation for strain GVR35 was determined by Bradford assay and was 461.15 µg ml^{-1}, while 427 was 451.97 µg ml^{-1} with a sample volume of 1 ml. To obtain a high enough protein concentration samples were precipitated with acetone and resuspended in 200µl DiGE lysis buffer. Final protein concentration was 600 µg ml^{-1} for strain GVR35 and 700 µg ml^{-1} for strain 427. Therefore, even after acetone precipitation a high sample volume had to be used to reach desired amount of protein (50 µg for DiGE gel). For the prep gels a lesser amount of sample had to be used; instead of 500 µg protein, only around 90 µg was loaded onto the prep gels. That resulted in a loss of protein spots showing in the prep gels as compared to the DiGE gel. Therefore not all protein spots showing differences could be analysed.

5.5.2. DiGE analysis

Protein spots with a minimum of twofold difference between strain GVR35 and strain 427 were picked for analysis. 35 protein spots were picked from the prep gel of strain GVR35 and 21 were picked from the strain 427 gel (Figure 5.1.). From the 35 spots picked from the GVR35 prep gel 28 proteins could be identified, while for strain 427, 17 of the 21 showed protein identification (Table 5.1. (427) and Table 5.2. (GVR35)).



(a)



(b)

Figure 5.1.: Image (DiGE) for *T. b. brucei* strain 427 (a) and GVR35 (b). 2-D separation was obtained using two step separation with an IEF strip pH 4-7 followed by SDS-Page size separation. Images were scanned and compared using DeCycler 2D software. Protein spots identified to show at least a twofold difference between the two strains were extracted from the prep gel, digested with trypsin and analysed by LC-MS. Protein spots extracted are marked in the images by a circle and the spot identification number for those were an identification could be made. The protein identifications can be seen in Table 5.1. for 427 and Table 5.2. for GVR35.

Spot No.	ID	Accession no.	Volume Ratio	Matches	Score
13	heat shock protein 83	Tb10.26.1080	-2.39	2(1)	37
54	69 kDa paraflagellar rod protein	Tb927.8.4970	-4.65	24(7)	186
374	hypothetical protein, conserved	Tb10.70.1130	-7.2	21(11)	313
502	protein phosphatase 2C, putative	Tb11.03.0390	-2.43	1(1)	38
539	arginine kinase	Tb09.160.4560	-2.37	2(1)	40
544	adenosine kinase, putative	Tb927.6.2360	-2.23	9(5)	119
549	arginine kinase	Tb09.160.4590	-9.33	46(17)	467
560	beta tubulin	Tb927.1.2330	-5.93	6(2)	71
571	asparagine synthetase a, putative	Tb927.7.1110	-5.66	4(0)	23
649	guanine nucleotide-binding protein beta subunit- like protein	Tb11.01.3170	-4.09	4(0)	28
677	nascent polypeptide associated complex subunit, putative	Tb09.211.0120	-2.37	4(1)	50
759	translation elongation factor 1-beta, putative	Tb927.4.3570	-7.27	27(13)	328
765	translation elongation factor 1-beta, putative	Tb10.70.1100	-7.24	26(12)	305
804	hypoxanthine-guanine phosphoribosyltransferase	Tb10.70.6540	-3.33	4(2)	66
888	profilin	Tb11.01.5350	-2.95	10(3)	139
891	60S acidic ribosomal protein, putative	Tb09.160.4200	-2.14	13(5)	164

Table 5.1.: Identified proteins from *T. b. brucei* strain 427 as extracted from gel (Figure 5.1. (a)). Spot no. can be matched to the location of the gel in Figure 5.1.(a). Identifications and accession no. were matched from mascot to TritrypDB. The volume ratio shows decreased/increased protein spot intensitiy and matches are the peptides matched against the protein sequence with the number in brackets hits indicating homology or identity. Score > 28 indicates homology or identity with $p > 0.05$.

Spot No.	ID	Accession no	Volume Ratio	Matches	Score
105	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	Tb10.6k15.2620	5.59	13(6)	192
111	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	Tb10.6k15.2620	4.74	16(5)	144
112	73 kDa paraflagellar rod protein	Tb927.3.4290	4.11	16(7)	199
176	S-adenosylhomocysteine hydrolase, putative	Tb11.01.1350	17.29	19(8)	198
199	RuvB-like DNA helicase, putative	Tb927.4.2000	8.69	4(1)	48
213	hypothetical protein, conserved	Tb927.7.3090	27.52	2(1)	40
246	malic enzyme, putative	Tb11.02.3130	7.14	2(1)	38
279	hypothetical protein, conserved	Tb927.7.3090	17.21	1(1)	41
286	variant surface glycoprotein (VSG, atypical), putative	Tb927.5.4950	9.76	1(1)	34
358	variant surface glycoprotein (VSG, pseudogene), putative	Tb927.6.5230	12.06		
371	enolase	Tb10.70.4740	3.05	29(17)	445
372	alpha tubulin	Tb927.1.2340	11.24	1(1)	33
417	heat shock protein 70	Tb11.01.3110	4.57	2(1)	54
498	alpha tubulin	Tb927.1.2340	17.61	1(1)	68
519	alpha tubulin	Tb927.1.2340	2.71	37(14)	406
527	alpha tubulin	Tb927.1.2340	21.49	2(1)	57
547	arginine kinase	Tb09.160.4560	8.19	6(2)	55
678	alpha tubulin	Tb927.1.2340	3.1	14(7)	218
694	alpha tubulin	Tb927.1.2340	32.44	1(1)	31
695	alpha tubulin	Tb927.1.2340	5.19	23(13)	434
701	beta tubulin	Tb927.1.2330	7.52	12(6)	193

Spot No.	ID	Accession no	Volume Ratio	Matches	Score
802	elongation factor 1 gamma, putative	Tb11.01.4660	4.25	6(1)	32
818	ADP-ribosylation factor-like protein 3A, putative	Tb927.3.3450	2.59	18(10)	216
862	translation elongation factor 1-beta, putative	Tb10.70.1100	7.27	8(3)	84
901	hypothetical protein, conserved	Tb10.26.0680	34.16	10(1)	62

Table 5.2.: Identified proteins from *T. b. brucei* strain GVR35 as extracted from gel (Figure 5.1. (b)). Spot no. can be matched to the location of the gel in Figure 5.1.(b). Identifications and accession no. were matched from mascot to TritrypDB. The volume ratio shows decreased/increased protein spot intensitiy and matches are the peptides matched against the protein sequence with the number in brackets hits indicating homology or identity. Score > 28 indicates homology or identity with $p > 0.05$.

The proteins identified in both strains included one hypothetical protein in 427 and two hypothetical proteins in GVR35.

Four proteins were identified in both datasets; however, comparing the 'protein spot' location in the gel (Figure 5.1) strongly indicates a difference between the two strains. Two isoforms of arginine kinase (AK) were extracted from 427 and one AK in GVR35. 427 contained AK1 and AK2, which are located in the flagellum and cytosol respectively, while in GVR35 only AK1 was detected. Heat shock protein 83, beta tubulin and translation elongation factor 1-beta (putative) were also detected in both datasets.

A first analysis of the detected proteins showed that most proteins were highly abundant in trypanosomes (M. Barrett, personal communication) and raised the question if the secretome samples might have been contaminated with proteins from lysed trypanosome cells. Although contamination from lysed cells cannot be excluded, a visual comparison between a trypanosome lysate 2-D gel and the two secretome 2-D gels showed differences in protein spot patterns and spot abundance. The secretome gel showed fewer spots and was biased towards low molecular weight proteins compared to a lysate gel (Dr R. Burchmore, personal communication).

To further investigate whether the contained proteins could be seen as a secretome sets, the obtained data were compared to already published trypanosome secretome sets.

Except for a small number of proteins in the two sets (profilin (427), RuvB-like DNA helicase, translation elongation factor 1-beta and malic enzyme (GVR35)), most proteins identified here have been described as being secreted from different strains of trypanosome before.

The role of profilin, identified in strain 427 dataset, in trypanosomes is unclear. It might be involved in the phosphoinositide signal transduction pathway in trypanosomes (Wilson and Seebeck, 1997). If the finding of the proteins RuvB-like DNA helicase, translation elongation factor 1-beta and malic enzyme in strain GVR35 are strain specific markers or just the result of contamination from dead cells still needs to be determined. Microscopical examination of the cells during incubation in serum free medium showed that GVR 35 was less tolerant to the conditions than 427.

RuvB-like DNA helicase, in complex with hsp 90, might be involved in cell proliferation, as suggested for leishmania and plasmodium (Ahmad et al., 2013) but the reason for this protein being potentially secreted is unclear.

To get a better overview of the whole set of secreted proteins in the sample set used, remaining sample was prepared for a gel free proteomics approach, using FASP (2.3.10).
The minimum amount of protein that can be analysed with this method is 5µg. For strain 427 14µg protein was used and for strain GVR35 14µg. Analysis of the secretome prepared by FASP showed only a small number of proteins detected for strain 427 and those samples were not further analysed. As very similar amounts of proteins were used and the samples were on the same LC-MS/MS run, a possible explanation for this is an experimental error using the FASP method. The most likely reason is that the proteins were not completely removed from the filter, but an error during LC-MS/MS run (For example sample not picked up properly) cannot be excluded.

5.2.2. *T. b. brucei* strain GVR35 secretome

After sample preparation and trypsin digest (as described in 2.3.10.), 109 proteins were significantly identified using MASCOT (p-value <0.05), with the identifications obtained from TritrypDB.

23 proteins were identified as hypothetical proteins and two unspecified products. The remaining 84 proteins had putative identifications belonging to different classes of proteins and were compared to published datasets (*T. congolense* (Grébaut et al., 2009) and *T. gambiense* (Geiger et al., 2010)). Proteins, detected in all three datasets (*T. brucei*, GVR35; *T. congolense* and *T. gambiense*) were divided into 10 groups (as seen in Figure 5.2.), namely Binding, Cytoskeleton organisation and flagellar proteins, Defence, Metabolism, Movement, Protein degradation, Protein folding, Protein synthesis, Signalling and Trafficking.

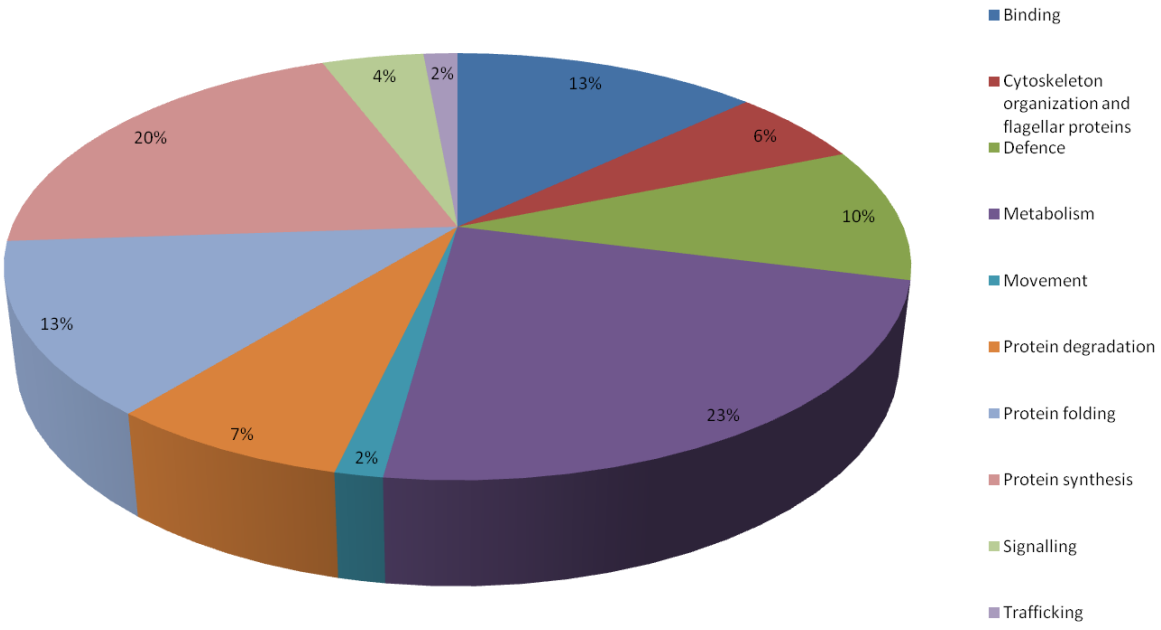


Figure 5.2.: Pie chart of classes of secreted proteins from *T.b.brucei* strain GVR35. 69 proteins were included in this analysis, as those proteins were also found in published secretome sets. Proteins were divided into classes, in accordance to published data. 23 % of the secreted proteins found in GVR 35 belonged to metabolism, representing the biggest group of secreted proteins. 20% of secreted proteins are involved in protein synthesis, the second biggest group. 13 % each belonging to Binding and Protein folding. Proteins belonging to the froup Defence, Protein degradation and Cytoskeleton organisiation and flagellar proteins were making 10%, 7% and 6% respectively. The rest (6% in total) are Signalling, Binding and Trafficking.

The largest group was represented by proteins involved in metabolism. This group can be divided into the following subgroups: Carbohydrate metabolism (9), amino acid metabolism (3), glutathione metabolism (2) and nucleic acid metabolism (2).

The presence of enzymes related to carbohydrate metabolism, also represent the largest protein group that the datasets had in common (6 out of 22 proteins).

Most enzymes of the glycolysis in trypanosomes are located in organelles called glycosomes, however, three glycolytic enzymes are described to be outside the glycosome (Enolase, phosphoglycerate mutase and pyruvate kinase; (Albert et al., 2005)) and those three enzymes were detected in all three compared datasets. A comparison between the GVR35 secretome and a published proteome of the glycosome (Colasante et al., 2006) showed 12 proteins in common (approximately 11% similarity between the two sets). The two sets shared five glycolytic enzymes namely fructose-bisphosphate aldolase, triosephosphate isomerase, glycerol kinase, glyceraldehyde 3-phosphate dehydrogenase and glycerol-3-phosphate dehydrogenase.

Enzymes belonging to the nucleic acid metabolism, in particular IAG nucleoside hydrolase, are important for bloodstream form trypanosomes, as they depend on the salvage pathway for purines (Parkin, 1996). Two were found in the GVR35 dataset (IAG nucleoside hydrolase and nucleoside diphosphate kinase), while in *T.b.gambiense* secretome the nucleotide metabolism proteins made up to 14% of the secreted proteins (Geiger et al., 2010).

The protein synthesis group is the second largest group. It is composed of ribosomal proteins and elongation factors. The role of these proteins in the secretome is unknown.

Proteins involved in protein folding and degradation have been previously described to have an effect on the immune system of the host, for example cyclophilin A and hsp (heat shock protein) (Calderwood et al., 2007; Kim et al., 2005), both being present in GVR35 secretome. Furthermore, it has been shown that cysteine peptidases play an important role in the pathogenesis of trypanosomes and it is suspected that it also helps the parasite to cross the blood brain barrier. However, *T.b.brucei* has been said to be less efficient than *T.b.gambiense* (Nikolskaia et al., 2006). The GVR35 dataset contained a calpain-like cysteine peptidase (Tb.927.7.4060) and metacaspase 4 (Tb10.70.5250). Metacaspase 4

(MCA4) belongs to the cysteine peptidases, but lacks peptidase activity (Proto et al., 2011). However, MCA4 knockout lines showed a reduced virulence in mice (Proto et al., 2011). Other proteins that reportedly have an impact on pathogenesis were calreticulin (Ferreira et al., 2004) which was not present in the GVR35 dataset and α/β-tubulin, who have shown to be T-cell stimulating antigens in leishmania infections (Probst et al., 2001). Both were detected in GVR35. Oligopeptidase B, an important virulence factor has been shown not to be secreted from cultured trypanosomes (Morty et al., 2001) and was also not detected in GVR35. A full list of all detected proteins can be found in Appendix E, Table E1.

Although the majority of proteins present in the GVR35 secretome set have also been described as being secreted from other trypanosome species, the high occurrence of glycolytic enzymes in the GVR35 secretome data was surprising and again raised the question if the presented dataset could be contaminated by proteins from lysed cells.

A western blot, using RAD51 (a cytosolic protein) as a control, showed minimal contamination in the secretome samples after 2 hours incubation (Figure 5.3). As cell viability during incubation in serum free medium was only checked using microscopic examination and strain GVR 35 showed cell clumps (sign for cell death) at the end of the incubation period, it cannot be ruled out that the secretome got contaminated with intra-cellular proteins.

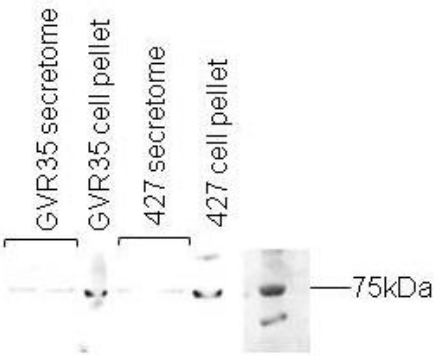


Figure 5.3: Western blot after secretome preparation, using cell pellet as control and Rad51 antibodies to check for contamination with dead cells.

5.2.3. FACS and cell survival

The original prepared and analysed secretome sets (by DiGE and FASP) indicated possible sample contamination from lysed parasites, due to the high number of glycolytic enzymes and highly abundant proteins present. To determine how many cells died during incubation in serum-free medium two different methods were tested: Fluorescent- activated cell sorting (FACS) using a MASCQuant and a propidium iodide (PI) assay based on (Gould et al., 2008).

FACS analysis was kindly done by S. Sabir (University of Glasgow), with propidium iodide used as a live/dead stain. The data showed that during incubation in serum free CMM about 5% of *T. brucei* strain 427 cells were measured as dead with the last measurement showing up to 8% dead cells. Strain GVR35 revealed about 18% cells as dead (Figure 5.4).

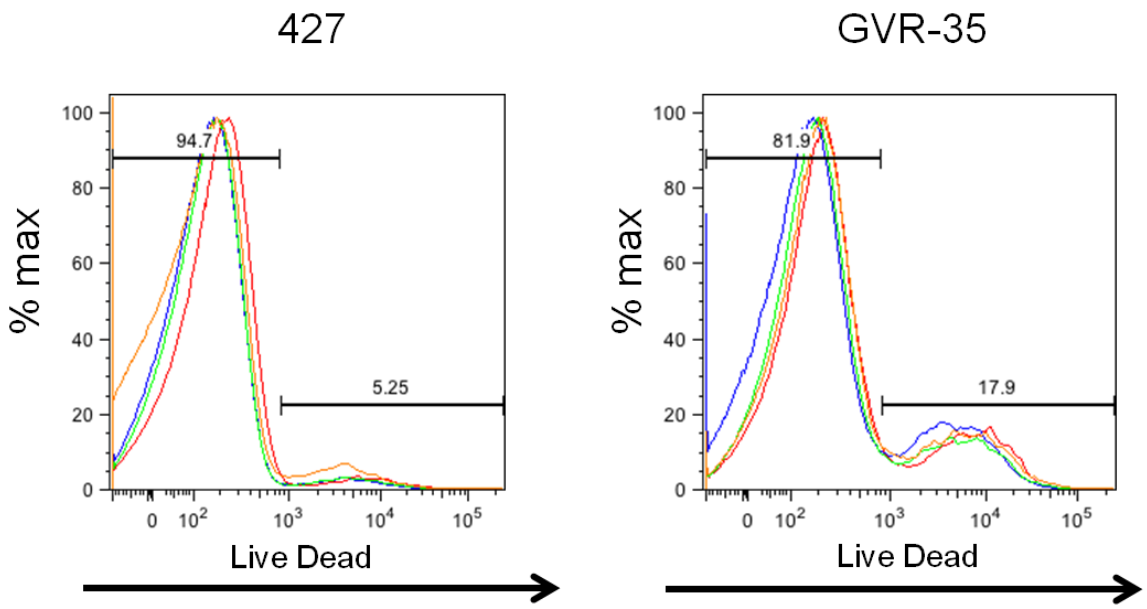


Figure 5.4.: FACS analysis on *T. b. brucei* strain 427wt and strain GVR35 during incubation in serum free CMM. Samples were taken every 30mins from beginning of incubation period. FACS analysis was kindly done by S. Sabir (University of Glasgow).

The propidium iodide assay was based on a method developed by Gould et al., who used it to monitor drug action in kinetoplastidae in real-time. For my purposes, cells of a known concentration were added to a 96 well plate and incubated with propidium iodide with measurements taken every 3 min. To determine the number of dead cells control samples of varying cell concentration were incubated with digitonin and propidium iodide. A standard curve with the fluorescent measurements taken from the dead cells was created, so the number of dead cells could be calculated (Figure 5.5).

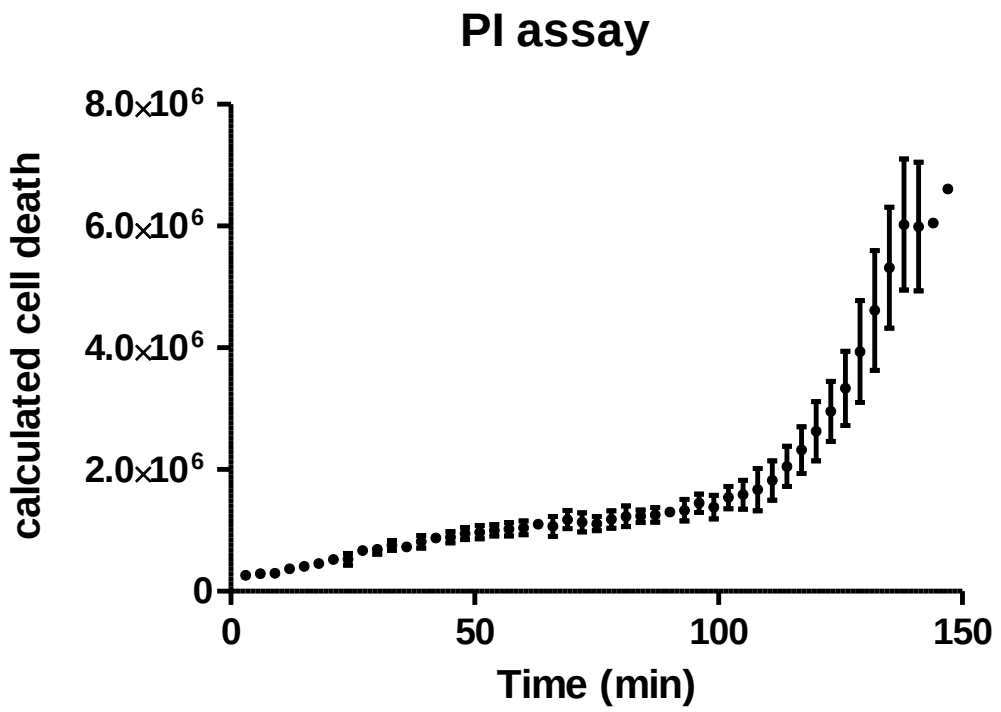


Figure 5.5: Real-time PI assay showing the increase of dead trypanosomes (427) during incubation in serum free CMM. Live/dead stain was achieved with propidium iodide and cells were incubated at a concentration of $5x10^7$. Four sample sets were used.

Cell death started to increase rapidly after 100 min. However, cell counts performed during cell incubation showed that cells kept on dividing even after being washed and kept in serum-free media. The analysis for the PI assay was performed for a static number of cells and as cell counts cannot be performed once the 96 well plate is set up for measurement this method does not seem accurate enough to determine the most accurate percentage of dead cells in the secretome. However, it did show the tolerance of the cells to serum free media, as shown by the rapid cell death after 100 minutes. As a consequence to this cell incubation time was reduced to 100 mins, however, results shown in 5.2.1 and 5.2.2 originated from sample incubated for 2 hours.

5.2.4. Dimethyl-labelling strain 427 and strain GVR35

As one aim was to compare the secretome of the strains 427 and GVR35, I made several attempts to repeat the DiGE secretome experiment. However, a protein concentration high enough to repeat the 2D-DiGE has not been achieved.

Dimethyl-labelling (Hsu et al., 2003) is a useful approach to get quantitative proteomics data with a smaller starting quantity of protein. Differentiation between two samples is achieved by labelling of the α- and ε-amino group of lysine residues via reductive amination. 11 µg of protein from each 427 and GVR35 were used, with 427 being heavy labelled and GVR35 light labelled.

Hit	Accession	Score	L/H	SD(geo)	Description
13	Tb09.211.3550	31	10.53	3.521	glycerol kinase, glycosomal
14	Tb09.211.3540	31	10.53	3.521	glycerol kinase, glycosomal
11	Tb11.01.3110	37	8.596	8.831	heat shock protein 70
17	Tb11.01.3080	26	8.596	8.831	heat shock protein 70, putative
28	Tb09.160.4560	16	6.108	1.593	arginine kinase
18	Tb11.03.0250	25	5.746	4.287	CYPA cyclophilin a; cyclophilin type peptidyl-prolyl cis-trans isomerase
2	**Tb10.v4.0053**	**116**	**4.562**	**1.051**	**hypothetical protein**
22	Tb10.70.5650	24	4.013	2.92	elongation factor 1-alpha
3	**Tb10.70.4740**	**97**	**3.552**	**1.346**	**enolase**
5	Tb10.70.5250	89	3.442	3.252	MCA4 metacaspase
10	**Tb10.70.1370**	**54**	**3.304**	**1.611**	**fructose-bisphosphate aldolase, glycosomal, putative**
1	**Tb927.1.2330**	**124**	**2.793**	**1.432**	**beta tubulin**

Table 5.3 Identified and quantified proteins from sample dimethyl-labelling. Proteins in bold have changed significantly between the two datasets. *T. b. brucei* strain GVR35 was light labelled and 427 was heavy labelled, significant changes in proteins are down in strain 427.

LC-MS/MS data analysed with Mascot distiller identified 30 proteins, 12 proteins with quantification of which 4 were significantly changed between datasets (Table 5.3).

The hypothetical protein Tb10.v4.0053 is listed as an unspecified product in TritrypDB for strain 427. For strain 927 and other trypanosome species and leishmania orthologs are identified as microtubule-associated protein. However, it needs to be taken as unidentified as not enough information exists to make a better annotation.

The enzyme enolase has been identified before in the 427 and GVR35 datasets. The DiGE approach showed enolase being increased in the GVR35 data and dimethyl-labelling confirmed this finding. Enolase, a glycolytic enzyme, catalyses the reversible conversion of D-2-phosphoglycerate and phosphoenolpyruvate. In trypanosomes, the enzyme is located outside the glycosomes and for leishmania can also been found bound to the cell surface where it displays no 'enolase' activity (Quiñones et al., 2007). However, enolase is also present on the cell surface of several mammalian cells and prokaryotic and eukaryotic pathogenic organisms (Avilán et al., 2011). It was found that enolase can act as a plasminogen-binding protein. The leishmania enolase seems to bind abound 60% of plasminogen to the cell surface (Vanegas et al., 2007). Plasminogen, a zymogen of serine protease plasmin, is part of the fibrinolytic system and it has been shown that pathogens can use this protease after binding it to their surface and activating it (Avilán et al., 2011).

Fructose-bisphosphate aldolase was detected in the GVR35 secretome data (5.2.2) and also in a *T. b. gambiense* secretome set (Geiger et al., 2010). However, a possible moonlighting function for this enzyme (similar to enolase) has not been described for trypanosomes.

Beta tubulin has a known association with the exosomes of various cell types (Olver and Vidal, 2007) and has also been identified in trypanosome secretome data.

5.3. Discussion

Proteomic approaches were used to identify proteins secreted from two different strains of *T. b. brucei* (namely strain 427 and strain GVR35) with the aim to establish (1) differences in secreted proteins between the two strains and (2) to obtain a profile of secreted proteins. Strain 427 and GVR35 show different pathogenesis in mice. While strain 427 causes an acute infection with high mortality within a few days, GVR35 establishes a chronic form of the disease with invasion of the CNS. Excreted/secreted proteins could be the key for a better understanding of the pathogenic process of those parasites, specifically in regards to invasion of the central nervous system.

This chapter introduced three different approaches to investigate the secretome of 427 and GVR35. However, the data is peliminary and does only indicate the possible set of secreted proteins. The rising cost of proteomic samples put a stop to this experiment before enough data was obtained to get comprehensive results.

5.3.1. Secretome production

Secretomes were produced by incubating the cells in serum free modified medium for two hours at a density of 2 x 10^8 cells ml^{-1} as described by Holzmueller et al. (2008). Cell viability was checked by microscopic examination throughout the incubation period. During the initial experiment (results 5.2.1 and 5.2.2) 427 cells showed good motility and minimal ‘clumping’. However, GVR35 showed less tolerance for these conditions and cell ‘clumping’ was observed during the end of the incubation period.

FACS analysis and a propidium iodide real-time assay on a repeat dataset confirmed that cells start to die after 100 min of incubation and a western blot analysis of secretome and cell lysate, using RAD51 as control, showed minimal contamination. FACS analysis showed about 18% dead cells in the GVR35 incubation culture compared to 5% in 427. That could explain the detected glycolytic enzymes in the analysed secretome sets from GVR35.

The propidium iodide assay indicated that cells were dying rapidly after 100 min. However, as cells kept on dividing in the serum free media (confirmed by cell counts) and

cell death was calculated assuming a static population this assay showed not to be accurate enough for the purpose of this experiment.

As a general observation, strain GVR35 seemed to cope less well with high density incubation than strain 427 and continuous cell cultures had to be kept between 1.5 and 2 x 10^6 cells ml^{-1} to avoid GVR35 parasite death (E. Myburgh, correspondence and personal experience). To resolve the issue of cell death during incubation several 'secretome productions' were performed at a lower density than the suggested 2 x 10^8 cells ml^{-1}, density tested were 1 x 10^8 and 5 x 10^7. Although the total number of cells was kept close to 2 x 10^8 used for the experiments, the protein amount gained was on average less than 100µg. However, lower cell density during incubation lowered cell death. A quantitative study on secreted proteins of *T. congolense* and *T.evansi* showed that not all species secrete proteins at the same level. *T. congolense* secreted five times less proteins than *T. evansi* under the same conditions (Holzmuller et al., 2008).

Secretome production has proven difficult to achieve in this study, but the reasons for this are unclear. It is possible, that the two strains used secrete lower level of proteins. Another possibility is the use of cultured cells. Published secretome compared to my own data had been isolated from rats and the interaction between host immune system and pathogen could trigger a higher amount of proteins being secreted. However, the classes of proteins detected in all sets were quite similar which suggest that trypanosomes do secrete proteins despite the lack of environmental pressure in culture. That leaves the amount of cells used for this study. Strain 427 and GVR35 grow well in cultures and 427 has been successfully grown in culture for decades. But the amount of culture that would be needed to achieve a high enough cell count to get a high enough amount of secreted protein would get unmanageable quite quickly. For strain GVR35, 150 to 170 ml of culture is needed to achieve 2 x 10^8 cells in total as cells start dying at densities over 2 x 10^6 (for continuous culture). From that amount on average 100µg of protein could be obtained. That means for DiGE analysis close to 1 litre of culture would be needed to get enough protein for one gel. Include three replicates and about three litres of culture becomes necessary. From a practical point of view alone this becomes unfeasible. For future experiments the use of trypanosomes isolated from mice might be a better approach to avoid the complications of cultured cells.

5.3.2. Comparison of *T. b. brucei* strains 427/GVR35

A comprehensive comparison between the strains 427 and GVR35 was not achieved due to the difficulties obtaining a high enough protein concentration for analysis with DiGE. Although some differences in protein presence was shown, the relatively small concentration of proteins in the secretome made it impossible to obtain enough protein to run the prep gels according to protocol and therefore made it difficult to analyse the data and pick spots from the prep gel. For the DiGE analysis protein spots showing an at least two-fold difference in ratio were extracted from the gel, digested with trypsin and analysed on a LC-MS/MS platform. Some proteins (for example alpha tubulin, GVR35 sample) were detected (or picked from the gel) several times as the two dimensional separation of the gel showed the protein in different positions (differences in charge and mass). However, this has been described for tubulins before (Bridges et al., 2008). Geiger et al showed modifications of proteins between different strains. This was evident from the DiGE approach between 427 and GVR35 as well, as four proteins were picked in both strains, but showed up as different spots on the DiGE gel.

A dimethyl-labelling approach identified 30 proteins but quantification was only possible for 12 proteins, with four proteins showing significant changes.

Both DiGE and dimethyl-labeling did show significant differences between the two strains. However, as experiments could not be repeated a comprehensive list of differential secreted proteins could not be obtained and further experiments are needed to support our hypothesis that secreted proteins are involved in the difference of virulence between the *T. b. brucei* strains 427 and GVR35.

5.3.3. *T. b. brucei* strain GVR 35 secretome

Although secretory processes in African trypanosomes have been studied and reviewed (Bangs, 1998; Clayton et al., 1995) certain aspects of the process still remains unclear. In trypanosomes, endocytosis and exocytosis occur through the flagellar pocket and require clathrin, actin and GTPase Rab proteins (Geiger et al., 2010). Clathrin heavy chain has been identified in the strain GVR35 dataset. Proteins associated with the exosome secretion pathway were identified and from the 22 proteins described (Olver and Vidal, 2007) 10 of them were present in this dataset, namely glyceraldehyde-3-phosphate

dehydrogenase, 14-3-3-like protein, alpha and beta tubulin, clathrin heavy chain, cyclophilin A, enolase, HSP70, pyruvate kinase and ubiquitin-conjugating enzyme.

A comparison between the strain GVR35 and other published secretome data has shown that from the 109 proteins identified 69 proteins were also found in other datasets. Secretome analysis from three different strains of *T. b. gambiense* revealed 50 % of secreted proteins belonging to the categories protein folding and degradation, nucleotide metabolism and unassigned function (Geiger et al., 2010). In comparison, in the two *T.b.brucei* strains tested here about 21% of the proteins were hypotheticals and proteins belonging to the category protein folding and degradation made up 20% of the analysed proteins in this dataset (see Figure 6.3). Only two enzymes belonging to the nucleic acid metabolism were detected, making it a significantly smaller proportion compared to the published datasets.

The proteins identified are connected to various molecular functions. Including, enzymes from carbohydrate metabolism, amino acid metabolism and nucleic acid metabolism, as well as chaperone proteins, proteins involved in protein and nucleotide binding, protein synthesis and cellular communication/signal transduction, which are categories known to be present in the secretome of trypanosomes.

Although some proteins are well described as being secreted by trypanosomes, others are not and finding evidence of them being secreted can lead to interesting new hypotheses about pathogenic role of secreted proteins. Geiger et al published in 2010 that three glycolysis related enzymes phosphoglycerate mutase, enolase, pyruvate kinase are found to be secreted (The strain GVR35 dataset contains all three). They suggest that those enzymes could have functions unrelated to glycolysis.

Comparison between the strain GVR35 and published secretome datasets showed a high number of enzymes from the carbohydrate metabolism, but if glycolysis related proteins in particular, are in fact secreted by trypanosomes or if their appearance is due to contamination of the secretome by cell death during incubation is unclear. However, glycolysis enzymes have been found in all published datasets for blood stream and procyclic (Atyame Nten et al., 2010) form of trypanosomes.

In conclusion, the comparison between my dataset and published datasets has shown that *T. b. brucei* strain GVR35 and to an extent strain 427, secrete similar proteins than other species of trypanosomes.

Chapter 6

Discussion

Whole genome sequencing has nowadays become more readily available with the cost dropping rapidly and data being made available for a variety of organisms, from bacteria to plants and trypanosomatids. The genome of several trypanosome species and subspecies has been published (Berriman et al., 2005; El-Sayed et al., 2005; Jackson et al., 2010) and data are available via databases, such as TritrypDB, a database dedicated to the trypanosomatids (Aslett et al., 2010). Gene annotation using bioinformatic approaches can give an indication of the possible function of a gene, but it is not guaranteed that the detected homology between two (or more) sequences actually means the proteins have functional identities. The most popular bioinformatics approaches being used are BLAST or Pfam. BLAST (Basic Local Alignment Search Tool) compares either nucleotide or protein sequences to sequences in databases. Pfam is used for protein annotation by searching a protein sequence for known protein domains (Punta et al., 2012). Although bioinformatics approaches for homology based functional assignments of genes has shown to be a fast approach to obtain putative gene annotations, often these annotations are incorrect or only based on low sequence homology (Baran et al., 2009). Identifying functions of hypothetical proteins or even identifying wrongly annotated genes remains a major challenge in the post-genomic era.

However, the field of metabolomics has created applications to overcome this challenge, which are (1) genes with no identified function and (2) metabolic pathways present but no encoded enzyme in the genome identified that could catalyse the reaction. However, metabolomics is the study of low molecular weight metabolites within the cell and metabolites are often the substrate or downstream product of enzymatic reactions, so seeing a reaction occur means there must be a gene present.

The first approach of directly identifying protein functions can be done by using recombinant protein or gene knock out / knock down. Analysing the changes in metabolic datasets can indicate the function of the protein of interest. Although this is the most direct approach, it is also time consuming.

A different approach is to look at the pathways directly. By combining stable isotope labelling with metabolomics, the flux of labelled compounds can be traced through the organism.

The metabolomic and proteomic approaches used in this study have shown to be highly applicable to study the system biology of trypanosomes.

Metabolomics and proteomics approaches have been applied to the parasitic protozoa *T. brucei* to (1) test a high throughput approach for enzyme function identification using metabolomics techniques, (2) identify metabolic pathways with stable isotope labelling coupled to metabolomics and (3) applying proteomic techniques for identification of novel secreted proteins from cultured parasites.

6.1. Enzyme assay / Enzyme ID

In this project, I investigated the potential of an untargeted enzyme assay using metabolomics techniques for a high throughput. This method has been deployed to identify the function of novel enzymes for *E. coli* (Saito et al., 2009) and *Mycobacterium tuberculosis* (De Cavallho et al., 2009; Larrouy-Maumus et al., 2013). Furthermore, three trypanosome enzymes have been investigated in our group (E. Kerkhoven, thesis).

In this project enzymes were chosen at random from a list of putative enzyme obtained from TritrypDB, as the aim was to determine whether this assay was applicable for a high throughput approach. The only criteria for the chosen 'enzymes' was that they should have a predicted metabolic function, relatively small size, so the cloning and over-expression would not cause too many difficulties, and that their function had not been determined before using recombinant or purified protein. The assay was designed to be fast and relatively cheap. The workflow developed took seven days to complete, from the initial PCR for cloning to purifying the protein for the assay. A way this was achieved was to use a ligase independent cloning system to standardise primer design and cut down on cloning time. When the system was applicable using *E. coli* over expression systems allowed easy and fast protein production. However, more than half of the proteins were either not over expressed or could not be purified (inclusion body formation), therefore alternative over expression systems should be considered for future work when over expression from *E. coli* is not possible. During this thesis, changes were made to optimise the workflow. Different *E. coli* over expression strains were tested to achieve high protein amounts. *E. coli* Rosetta (DE3) cells improved protein expression in number of proteins that could be

over expressed and their yield compared to BL21 (DE3) strains. A second extraction step was also added to the protocol to increase the quality of the metabolomics datasets (as was shown with the example of Hexokinase, Chapter 3). The concentration of the MOPS buffer used in the enzyme assay was lowered to 10 mM from originally 40 mM, to reduce the possibility of ion suppression and also because MOPS was seen to block the columns.

For an initial screening approach, this method seems applicable as shown for the enzyme S-adenosylmethionine synthetase. Using the *in vitro* assay the function of this enzyme was directly linked to a purified protein. The presence of this reaction was previously achieved using trypanosome cell extract (Yarlett et al., 1993), but the function was not linked directly to a gene.

But not all proteins showed changes indicating its function, highlighting the limitations of this method.

Currently there is no MS platform fitting for all classes of metabolites, which causes a problem when the substrate/products fall outwith the parameters suitable for analysis on the LC-MS platform used in this project. For putative enzymes in this study it is hard to tell if that is the case, however, the commercial hexokinase, used for validation purposes shows that detection of sugars is not ideal on LC-MS, as the separation is not good enough to differential between glucose and fructose (for example). Most sugars and sugar phosphates are being identified with 57 isomers on the IDEOM spreadsheet.

Yeast extract has shown to be a broad, reproducible source for metabolites, which has the benefit, in the high throughput approach, to limit the external variance factor. The drawback is that trypanosome specific enzymes might not be identified as specific metabolites, like trypanothione, might be missing. Also, highly abundant metabolites could mask changes, as was shown for S-adenosylmethionine synthetase, where levels of L-methionine were unchanged between the treatment and control samples and for the commercial hexokinase assay, which showed no changes in D-glucose levels.

One of the proteins investigated, putatively annotated as deoxhypusine synthase, was since shown to form a heterotetramer with a catalytically dead paralog to enhance its activity by 3000-fold (Nguyen et al., 2013). This form of regulation, termed prozyme, has been described for other trypanosome enzymes as well, namely hexokinase (Morris et al., 2006)

and S-adenosylmethionine decarboxylase (Velez et al., 2013). The use of the untargeted enzyme assay with recombinant proteins will miss those enzymes as well.

When the function of a potential enzyme cannot be determined using the *in vitro* assay approach, the production of knock-out lines or knock-down lines can be of use to further investigate the function (Saghatellian et al., 2004). However, this process can be very time consuming compared to the production of recombinant protein in a well established over expression system such as *E. coli*. The benefit of this approach, next to determine the function of the targeted gene, is to also provide further information about the target protein, such as how essential this protein is for cell survival.

Metabolomics analysis of knock out or knock down lines does not always allow identification of the function. For example, the function of the trypanosome 'arginase' has eluded our group for years now. This protein has been investigated in three PhD studies and all we can say about it is that it does not have 'arginase' activity although BLAST and Pfam show significant homologies to other arginases. An 'arginase' knock-out mutant was produced and using an untargeted metabolomics approach comparing knock-out to wildtype gave no indication of arginase or ureohydrolase activity (Hai et al., 2014).

6.2. Stable Isotope Labelling / Pathway ID

Metabolomics combined with stable isotope labelling can provide an indirect approach for the detection of enzymatic activity (Dalluge et al., 2005). However, this approach is mainly used as a tool for pathway identifications, as by adding a labelled metabolite to the cell culture, downstream metabolites labelled from the added compound can be detected and lead to the identification of novel pathways. Extensive work has been done on the energy metabolism of procyclic and bloodstream form trypanosome. Recent work from Creek et al (2015) used U- ^{13}C D-glucose to label cultured bloodstream form trypanosomes, showing that glucose can enter many branches of the trypanosome metabolism.

ATP, labelled from U-^{13}C glucose, was shown to be incorporated in S-adenosylmethionine. Trypanosomes were thought to be capable of recycling L-methionine via the methylthioadenosine cycle (MTA cycle or Yang cycle). However, L-methionine was not detected to be labelled from glucose, ruling out the use of L-methionine salvage via the

MTA cycle. To further investigate those findings, U-^{13}C-L-methionine was used in an untargeted metabolomic approach.

The main findings are:

- The MTA (or Yang) cycle also does not seem to be involved in methionine salvage. Instead MTR gets secreted from the cells, shown by rising levels of 1-C labelled MTR in spent medium. Also, the precursor metabolite to L-methionine (4-methylthio-2-oxobutanoate) is 5-^{13}C labelled from L-methionine, confirming the findings of Creek et al., 2015.
- 1,2-dihydroxy-5-(methylthio) pent-1-en-3-one from the MTA cycle is labelled 1-^{13}C, but no further information about what pathway this metabolite goes into was obtained.
- The Methionine cycle, recycling S-adenosylmethionine to L-homocysteine back to L-methionine doesn't seem to take place (lack of 1-C label in L-methionine)

Future works, could include the detection of enzymes of the MTA cycle present in trypanosomes and further investigate the fate of 1,2-dihydroxy-5-(methylthio) pent-1-en-3-one in bloodstream form trypanosomes.

Another finding of the U-^{13}C glucose study (Creek et al., 2015) was a significant level of aspartate, succinate and malate not made from glucose, suggesting alternative carbon source than glucose for their production. Data obtained from U-^{13}C L-glutamine labelling in cultured trypanosomes (DH Kim and F Achcar, unpublished data) showed labelled carbon incorporation from L-glutamine. However, approximately 20% of aspartate, succinate and malate were still unaccounted for. As L-proline can be the main carbon source in procyclic trypanosomes, U-^{13}C L-proline was used to investigate its possible involvement in bloodstream form trypanosomes. However, minimal labelling of compounds in the dataset suggests L-proline is not used as carbon source in bloodstream form trypanosomes. No carbon incorporation into those compounds was detected from L-methionine and L-arginine. Further testing of other amino acids is required to answer for the unaccounted amounts of aspartate, succinate and malate. Enzymes for the oxidation of aromatic amino acids are absent from trypanosomes (Berriman *et al*., 2005) making their involvement unlikely. L-aspartate is not taken up by trypanosomes (Hasne, thesis) and work done by D Kim with ^{13}C- L-cysteine (unpublished data) showed no labelling.

Previous work from I. Vincent on the L-ornithine biosynthesis showed that trypanosomes take up L-ornithine from the media. However, when cells were incubated in CBSS (Carter's balanced saline solution) in the absence of L-ornithine and with added ^{14}N-L-arginine, labelling of L-ornithine occurred. The most direct route from L-arginine to L-ornithine is via the enzyme arginase. *T. b. brucei'*s genome encodes for a gene annotated as arginase (or agmatinase-like protein), however, this enzymes has no arginase activity (Hai *et al*., 2014). To investigate an alternative route for L-ornithine production, but also to further probe the L-arginine metabolism in trypanosomes, U-^{13}C-L-arginine was used in an untargeted metabolomic approach.

The main findings are:

- Although L-ornithine was 5-C labelled from L-arginine, equally high concentrations were seen intra- and extracellular (spent medium analysis). Therefore, it was assumed that the formation of L-ornithine happened outside the cell and labelled L-ornithine was taken up by the trypanosomes. This was later verified by Dr. F. Giordani by incubating medium with added U-C^{13} L-arginine without cells.
- Arginine-phosphate is labelled from L-arginine. This has been described to be of use in energy storage in trypanosomes (Canepa et al., 2011), with three arginine kinases being encoded in the *T. brucei* genome.

6.3. Secretome

Proteomic techniques were used to identify secreted proteins from two different trypanosome strains, which differ in the course of disease, 427 causes an acute infection in mice while GVR35 remains as a more chronic form within the CNS of the mice. In this thesis, I showed significant differences in proteins secreted from strain 427 and GVR35, although the results are just preliminary. However, the preliminary data did show a high consensus between the possible secreted proteins from strain 427 and GVR35 to other published secretome datasets, including one from human infecting trypanosomes (Geiger et al., 2010).

Although published trypanosome secretome datasets included 'proof' of minimal contamination of proteins from lysed cells by FACS and western blot analysis, the

occurrence of highly abundant proteins and glycolytic enzymes raises the questions if those proteins are actively secreted or not.

Further work is needed to address this problem and for the strains used in this study alternative methods for secretome production are needed.

List of References

Abdulla, M.-H., O'Brien, T., Mackey, Z.B., Sajid, M., Grab, D.J., and McKerrow, J.H. (2008). RNA Interference of Trypanosoma brucei Cathepsin B and L Affects Disease Progression in a Mouse Model. PLoS Negl. Trop. Dis. *2*.

Ahmad, M., Afrin, F., and Tuteja, R. (2013). Identification of R2TP complex of Leishmania donovaniand Plasmodium falciparum using genome wide in-silico analysis. Commun. Integr. Biol. *6*.

Albert, M.-A., Haanstra, J.R., Hannaert, V., Van Roy, J., Opperdoes, F.R., Bakker, B.M., and Michels, P.A.M. (2005). Experimental and in silico analyses of glycolytic flux control in bloodstream form Trypanosoma brucei. J. Biol. Chem. *280*, 28306–28315.

Alsford, S., and Horn, D. (2008). Single-locus targeting constructs for reliable regulated RNAi and transgene expression in Trypanosoma brucei. Mol. Biochem. Parasitol. *161*, 76–79.

Alves, M.J.M., and Colli, W. (2008). Role of the gp85/trans-sialidase superfamily of glycoproteins in the interaction of Trypanosoma cruzi with host structures. Subcell. Biochem. *47*, 58–69.

Aman, R.A., Kenyon, G.L., and Wang, C.C. (1985). Cross-linking of the enzymes in the glycosome of Trypanosoma brucei. J. Biol. Chem. *260*, 6966–6973.

Antelmann, H., Tjalsma, H., Voigt, B., Ohlmeier, S., Bron, S., van Dijl, J.M., and Hecker, M. (2001). A proteomic view on genome-based signal peptide predictions. Genome Res. *11*, 1484–1502.

Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B.P., Carrington, M., Depledge, D.P., Fischer, S., Gajria, B., Gao, X., et al. (2010). TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res. *38*, D457–D462.

Atyame Nten, C.M., Sommerer, N., Rofidal, V., Hirtz, C., Rossignol, M., Cuny, G., Peltier, J.-B., and Geiger, A. (2010). Excreted/secreted proteins from trypanosome procyclic strains. J. Biomed. Biotechnol. *2010*, 212817.

Avilán, L., Gualdrón-López, M., Quiñones, W., González-González, L., Hannaert, V., Michels, P.A.M., and Concepción, J.-L. (2011). Enolase: a key player in the metabolism and a probable virulence factor of trypanosomatid parasites-perspectives for its use as a therapeutic target. Enzyme Res. *2011*, 932549.

Bacchi, C.J., and Yarlett, N. (1993). Effects of antagonists of polyamine metabolism on African trypanosomes. Acta Trop. *54*, 225–236.

Bacchi, C.J., Goldberg, B., Garofalo-Hannan, J., Rattendi, D., Lyte, P., and Yarlett, N. (1995). Fate of soluble methionine in African trypanosomes: effects of metabolic inhibitors. Biochem. J. *309*, 737–743.

Bangs, J.D. (1998). Surface coats and secretory trafficking in African trypanosomes. Curr. Opin. Microbiol. *1*, 448–454.

Bangs, J.D., Brouch, E.M., Ransom, D.M., and Roggy, J.L. (1996). A Soluble Secretory Reporter System in Trypanosoma brucei STUDIES ON ENDOPLASMIC RETICULUM TARGETING. J. Biol. Chem. *271*, 18387–18393.

Baran, R., Reindl, W., and Northen, T.R. (2009). Mass spectrometry based metabolomics and enzymatic assays for functional genomics. Curr. Opin. Microbiol. *12*, 547–552.

Barrett, M.P. (2006). The rise and fall of sleeping sickness. Lancet *367*, 1377–1378.

Barrett, M.P., Burchmore, R.J.S., Stich, A., Lazzari, J.O., Frasch, A.C., Cazzulo, J.J., and Krishna, S. (2003). The trypanosomiases. Lancet *362*, 1469–1480.

Barrett, M.P., Boykin, D.W., Brun, R., and Tidwell, R.R. (2007). Human African trypanosomiasis: pharmacological re-engagement with a neglected disease. Br. J. Pharmacol. *152*, 1155–1171.

Barrett, M.P., Bakker, B.M., and Breitling, R. (2010). Metabolomic systems biology of trypanosomes. Parasitology *137*, 1285–1290.

Beckonert, O., Keun, H.C., Ebbels, T.M.D., Bundy, J., Holmes, E., Lindon, J.C., and Nicholson, J.K. (2007). Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. Nat. Protoc. *2*, 2692–2703.

Benne, R., Van Den Burg, J., Brakenhoff, J.P.J., Sloof, P., Van Boom, J.H., and Tromp, M.C. (1986). Major transcript of the frameshifted coxll gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell *46*, 819–826.

Berger, B.J., Dai, W.W., Wang, H., Stark, R.E., and Cerami, A. (1996). Aromatic amino acid transamination and methionine recycling in trypanosomatids. Proc. Natl. Acad. Sci. U. S. A. *93*, 4126–4130.

Berger, L.C., Wilson, J., Wood, P., and Berger, B.J. (2001). Methionine regeneration and aspartate aminotransferase in parasitic protozoa. J. Bacteriol. *183*, 4421–4434.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D.C., Lennard, N.J., Caler, E., Hamlin, N.E., Haas, B., et al. (2005). The genome of the African trypanosome Trypanosoma brucei. Science *309*, 416–422.

Bitonti, A.J., Dumont, J.A., and McCann, P.P. (1986). Characterization of Trypanosoma brucei brucei S-adenosyl-L-methionine decarboxylase and its inhibition by Berenil, pentamidine and methylglyoxal bis(guanylhydrazone). Biochem. J. *237*, 685–689.

Blum, J., and Burri, C. (2002). Treatment of late stage sleeping sickness caused by T.b. gambiense: a new approach to the use of an old drug. Swiss Med. Wkly. *132*, 51–56.

Blum, J.A., Zellweger, M.J., Burri, C., and Hatz, C. (2008). Cardiac involvement in African and American trypanosomiasis. Lancet Infect. Dis. *8*, 631–641.

Breitling, R., Pitt, A.R., and Barrett, M.P. (2006a). Precision mapping of the metabolome. Trends Biotechnol. *24*, 543–548.

Breitling, R., Ritchie, S., Goodenowe, D., Stewart, M.L., and Barrett, M.P. (2006b). Ab initio prediction of metabolic networks using Fourier transform mass spectrometry data. Metabolomics *2*, 155–164.

Bridges, D.J., Pitt, A.R., Hanrahan, O., Brennan, K., Voorheis, H.P., Herzyk, P., de Koning, H.P., and Burchmore, R.J.S. (2008). Characterisation of the plasma membrane subproteome of bloodstream form Trypanosoma brucei. Proteomics *8*, 83–99.

Bringaud, F., Rivière, L., and Coustou, V. (2006). Energy metabolism of trypanosomatids: adaptation to available carbon sources. Mol. Biochem. Parasitol. *149*, 1–9.

Brown, M., Wedge, D.C., Goodacre, R., Kell, D.B., Baker, P.N., Kenny, L.C., Mamas, M.A., Neyses, L., and Dunn, W.B. (2011). Automated workflows for accurate mass-based putative metabolite identification in LC/MS-derived metabolomic datasets. Bioinformatics *27*, 1108–1112.

Brun, R., Blum, J., Chappuis, F., and Burri, C. (2010). Human African trypanosomiasis. Lancet *375*, 148–159.

Burkard, G., Fragoso, C.M., and Roditi, I. (2007). Highly efficient stable transformation of bloodstream forms of Trypanosoma brucei. Mol. Biochem. Parasitol. *153*, 220–223.

Bursell, E. (1963). Aspects of the metabolism of amino acids in the tsetse fly, Glossina (Diptera). J. Insect Physiol. *9*, 439–452.

Calderwood, S.K., Mambula, S.S., Gray, P.J., and Theriault, J.R. (2007). Extracellular heat shock proteins in cell signaling. FEBS Lett. *581*, 3689–3694.

Canepa, G.E., Carrillo, C., Miranda, M.R., Sayé, M., and Pereira, C.A. (2011). Arginine kinase in Phytomonas, a trypanosomatid parasite of plants. Comp. Biochem. Physiol. B Biochem. Mol. Biol. *160*, 40–43.

De Carvalho, L.P.S., Zhao, H., Dickinson, C.E., Arango, N.M., Lima, C.D., Fischer, S.M., Ouerfelli, O., Nathan, C., and Rhee, K.Y. (2010). Activity-based metabolomic profiling of enzymatic function: identification of Rv1248c as a mycobacterial 2-hydroxy-3-oxoadipate synthase. Chem. Biol. *17*, 323–332.

Castle, A.L., Fiehn, O., Kaddurah-Daouk, R., and Lindon, J.C. (2006). Metabolomics Standards Workshop and the development of international standards for reporting metabolomics experimental results. Brief. Bioinform. *7*, 159–165.

Cazzulo, J.J. (1992). Aerobic fermentation of glucose by trypanosomatids. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. *6*, 3153–3161.

Chambers, J.W., Kearns, M.T., Morris, M.T., and Morris, J.C. (2008). Assembly of heterohexameric trypanosome hexokinases reveals that hexokinase 2 is a regulable enzyme. J. Biol. Chem. *283*, 14963–14970.

Choi, J., and El-Sayed, N.M. (2012). Functional genomics of trypanosomatids. Parasite Immunol. *34*, 72–79.

Chokkathukalam, A., Jankevics, A., Creek, D.J., Achcar, F., Barrett, M.P., and Breitling, R. (2013). mzMatch-ISO: an R tool for the annotation and relative quantification of isotope-labelled mass spectrometry data. Bioinforma. Oxf. Engl. *29*, 281–283.

Clayton, C., Häusler, T., and Blattner, J. (1995). Protein trafficking in kinetoplastid protozoa. Microbiol. Rev. *59*, 325–344.

Colasante, C., Ellis, M., Ruppert, T., and Voncken, F. (2006). Comparative proteomics of glycosomes from bloodstream form and procyclic culture form Trypanosoma brucei brucei. Proteomics *6*, 3275–3293.

Coustou, V., Biran, M., Breton, M., Guegan, F., Rivière, L., Plazolles, N., Nolan, D., Barrett, M.P., Franconi, J.-M., and Bringaud, F. (2008). Glucose-induced remodeling of intermediary and energy metabolism in procyclic Trypanosoma brucei. J. Biol. Chem. *283*, 16342–16354.

Creek, D.J., Jankevics, A., Breitling, R., Watson, D.G., Barrett, M.P., and Burgess, K.E.V. (2011). Toward global metabolomics analysis with hydrophilic interaction liquid chromatography-mass spectrometry: improved metabolite identification by retention time prediction. Anal. Chem. *83*, 8703–8710.

Creek, D.J., Anderson, J., McConville, M.J., and Barrett, M.P. (2012a). Metabolomic analysis of trypanosomatid protozoa. Mol. Biochem. Parasitol. *181*, 73–84.

Creek, D.J., Jankevics, A., Burgess, K.E.V., Breitling, R., and Barrett, M.P. (2012b). IDEOM: an Excel interface for analysis of LC-MS-based metabolomics data. Bioinforma. Oxf. Engl. *28*, 1048–1049.

Creek, D.J., Chokkathukalam, A., Jankevics, A., Burgess, K.E.V., Breitling, R., and Barrett, M.P. (2012c). Stable isotope-assisted metabolomics for network-wide metabolic pathway elucidation. Anal. Chem. *84*, 8442–8447.

Creek, D.J., Nijagal, B., Kim, D.-H., Rojas, F., Matthews, K.R., and Barrett, M.P. (2013). Metabolomics guides rational development of a simplified cell culture medium for drug screening against Trypanosoma brucei. Antimicrob. Agents Chemother. *57*, 2768–2779.

Creek, D.J., Mazet, M., Achcar, F., Anderson, J., Kim, D.-H., Kamour, R., Morand, P., Millerioux, Y., Biran, M., Kerkhoven, E.J., et al. (2015). Probing the Metabolic Network in Bloodstream-Form Trypanosoma brucei Using Untargeted Metabolomics with Stable Isotope Labelled Glucose. PLoS Pathog. *11*, e1004689.

Cross, G. a. M. (1975). Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of Trypanosoma brucei. Parasitology *71*, 393–417.

Cubbon, S., Antonio, C., Wilson, J., and Thomas-Oates, J. (2010). Metabolomic applications of HILIC-LC-MS. Mass Spectrom. Rev. *29*, 671–684.

Dai, Y., Pochapsky, T.C., and Abeles, R.H. (2001). Mechanistic studies of two dioxygenases in the methionine salvage pathway of Klebsiella pneumoniae. Biochemistry (Mosc.) *40*, 6379–6387.

Dalluge, J.J., Liao, H., Gokarn, R., and Jessen, H. (2005). Discovery of enzymatic activity using stable isotope metabolite labeling and liquid chromatography-mass spectrometry. Anal. Chem. *77*, 6737–6740.

Dettmer, K., Aronov, P.A., and Hammock, B.D. (2007). Mass spectrometry-based metabolomics. Mass Spectrom. Rev. *26*, 51–78.

Dumon-Seignovert, L., Cariot, G., and Vuillard, L. (2004). The toxicity of recombinant proteins in Escherichia coli: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). Protein Expr. Purif. *37*, 203–206.

Dunn, W.B. (2008). Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. Phys. Biol. *5*, 011001.

Dunn, W.B., and Ellis, D.I. (2005). Metabolomics: Current analytical platforms and methodologies. TrAC Trends Anal. Chem. *24*, 285–294.

Dunn, W.B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., Brown, M., Knowles, J.D., Halsall, A., Haselden, J.N., et al. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nat. Protoc. *6*, 1060–1083.

Duszenko, M., Ivanov, I.E., Ferguson, M.A., Plesken, H., and Cross, G.A. (1988). Intracellular transport of a variant surface glycoprotein in Trypanosoma brucei. J. Cell Biol. *106*, 77–86.

Ebikeme, C., Hubert, J., Biran, M., Gouspillou, G., Morand, P., Plazolles, N., Guegan, F., Diolez, P., Franconi, J.-M., Portais, J.-C., et al. (2010). Ablation of succinate production from glucose metabolism in the procyclic trypanosomes induces metabolic switches to the glycerol 3-phosphate/dihydroxyacetone phosphate shuttle and to proline metabolism. J. Biol. Chem. *285*, 32312–32324.

El-Sayed, N.M., Myler, P.J., Bartholomeu, D.C., Nilsson, D., Aggarwal, G., Tran, A.-N., Ghedin, E., Worthey, E.A., Delcher, A.L., Blandin, G., et al. (2005). The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. Science *309*, 409–415.

Englund, P.T., Hajduk, S.L., and Marini, J.C. (1982). The Molecular Biology of Trypanosomes. Annu. Rev. Biochem. *51*, 695–726.

Engstler, M., Thilo, L., Weise, F., Grünfelder, C.G., Schwarz, H., Boshart, M., and Overath, P. (2004). Kinetics of endocytosis and recycling of the GPI-anchored variant surface glycoprotein in Trypanosoma brucei. J. Cell Sci. *117*, 1105–1115.

Engstler, M., Pfohl, T., Herminghaus, S., Boshart, M., Wiegertjes, G., Heddergott, N., and Overath, P. (2007). Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes. Cell *131*, 505–515.

Enyaru, J.C.K., Matovu, E., Nerima, B., Akol, M., and Sebikali, C. (2006). Detection of T.b. rhodesiense trypanosomes in humans and domestic animals in south east Uganda by amplification of serum resistance-associated gene. Ann. N. Y. Acad. Sci. *1081*, 311–319.

Fairlamb, A.H. (1989). Novel biochemical pathways in parasitic protozoa. Parasitology *99 Suppl*, S93–S112.

Fairlamb, A.H., and Cerami, A. (1992). Metabolism and functions of trypanothione in the Kinetoplastida. Annu. Rev. Microbiol. *46*, 695–729.

Fairlamb, A.H., Blackburn, P., Ulrich, P., Chait, B.T., and Cerami, A. (1985). Trypanothione: a novel bis(glutathionyl)spermidine cofactor for glutathione reductase in trypanosomatids. Science *227*, 1485–1487.

Fairlamb, A.H., Henderson, G.B., and Cerami, A. (1989). Trypanothione is the primary target for arsenical drugs against African trypanosomes. Proc. Natl. Acad. Sci. U. S. A. *86*, 2607–2611.

Fan, T.W.-M., Lorkiewicz, P., Sellers, K., Moseley, H.N.B., Higashi, R.M., and Lane, A.N. (2012). Stable isotope-resolved metabolomics and applications for drug development. Pharmacol. Ther. *133*, 366–391.

Ferguson, M.A., and Cross, G.A. (1984). Myristylation of the membrane form of a Trypanosoma brucei variant surface glycoprotein. J. Biol. Chem. *259*, 3011–3015.

Ferguson, M.A., and Williams, A.F. (1988). Cell-surface anchoring of proteins via glycosyl-phosphatidylinositol structures. Annu. Rev. Biochem. *57*, 285–320.

Field, M.C., and Carrington, M. (2004). Intracellular membrane transport systems in Trypanosoma brucei. Traffic Cph. Den. *5*, 905–913.

Fisk, J.C., Li, J., Wang, H., Aletta, J.M., Qu, J., and Read, L.K. (2013). Proteomic analysis reveals diverse classes of arginine methylproteins in mitochondria of trypanosomes. Mol. Cell. Proteomics MCP *12*, 302–311.

Franco, J.R., Simarro, P.P., Diarra, A., and Jannin, J.G. (2014). Epidemiology of human African trypanosomiasis. Clin. Epidemiol. *6*, 257–275.

Frevert, U., Movila, A., Nikolskaia, O.V., Raper, J., Mackey, Z.B., Abdulla, M., McKerrow, J., and Grab, D.J. (2012). Early Invasion of Brain Parenchyma by African Trypanosomes. PLoS ONE *7*.

Fridman, E., and Pichersky, E. (2005). Metabolomics, genomics, proteomics, and the identification of enzymes and their substrates and products. Curr. Opin. Plant Biol. *8*, 242–248.

Garcia, A., and Barbas, C. (2011). Gas chromatography-mass spectrometry (GC-MS)-based metabolomics. Methods Mol. Biol. Clifton NJ *708*, 191–204.

Garcia, D.E., Baidoo, E.E., Benke, P.I., Pingitore, F., Tang, Y.J., Villa, S., and Keasling, J.D. (2008). Separation and mass spectrometry in microbial metabolomics. Curr. Opin. Microbiol. *11*, 233–239.

Garzón, E., Holzmuller, P., Bras-Gonçalves, R., Vincendeau, P., Cuny, G., Lemesre, J.L., and Geiger, A. (2013). The Trypanosoma brucei gambiense secretome impairs lipopolysaccharide-induced maturation, cytokine production, and allostimulatory capacity of dendritic cells. Infect. Immun. *81*, 3300–3308.

Geiger, A., Hirtz, C., Bécue, T., Bellard, E., Centeno, D., Gargani, D., Rossignol, M., Cuny, G., and Peltier, J.-B. (2010). Exocytosis and protein secretion in Trypanosoma. BMC Microbiol. *10*, 20.

Ghoda, L.Y., Savarese, T.M., Northup, C.H., Parks, R.E., Garofalo, J., Katz, L., Ellenbogen, B.B., and Bacchi, C.J. (1988). Substrate specificities of 5'-deoxy-5'-methylthioadenosine phosphorylase from Trypanosoma brucei brucei and mammalian cells. Mol. Biochem. Parasitol. *27*, 109–118.

Gibson, W. (2007). Resolution of the species problem in African trypanosomes. Int. J. Parasitol. *37*, 829–838.

Gonçalves, M.F., Umezawa, E.S., Katzin, A.M., de Souza, W., Alves, M.J., Zingales, B., and Colli, W. (1991). Trypanosoma cruzi: shedding of surface antigens as membrane vesicles. Exp. Parasitol. *72*, 43–53.

Gould, M.K., Vu, X.L., Seebeck, T., and de Koning, H.P. (2008). Propidium iodide-based methods for monitoring drug action in the kinetoplastidae: comparison with the Alamar Blue assay. Anal. Biochem. *382*, 87–93.

Grébaut, P., Chuchana, P., Brizard, J.-P., Demettre, E., Seveno, M., Bossard, G., Jouin, P., Vincendeau, P., Bengaly, Z., Boulangé, A., et al. (2009). Identification of total and differentially expressed excreted-secreted proteins from Trypanosoma congolense strains exhibiting different virulence and pathogenicity. Int. J. Parasitol. *39*, 1137–1150.

Grønborg, M., Kristiansen, T.Z., Iwahori, A., Chang, R., Reddy, R., Sato, N., Molina, H., Jensen, O.N., Hruban, R.H., Goggins, M.G., et al. (2006). Biomarker discovery from pancreatic cancer secretome using a differential proteomic approach. Mol. Cell. Proteomics MCP *5*, 157–171.

Haggarty, J., Oppermann, M., Dalby, M.J., Burchmore, R.J., Cook, K., Weidt, S., and Burgess, K.E.V. (2015). Serially coupling hydrophobic interaction and reversed-phase chromatography with simultaneous gradients provides greater coverage of the metabolome. Metabolomics 1–6.

Hai, Y., Kerkhoven, E.J., Barrett, M.P., and Christianson, D.W. (2015). Crystal structure of an arginase-like protein from Trypanosoma brucei that evolved without a binuclear manganese cluster. Biochemistry (Mosc.) *54*, 458–471.

Hajduk, S.L., Hager, K.M., and Esko, J.D. (1994). Human high density lipoprotein killing of African trypanosomes. Annu. Rev. Microbiol. *48*, 139–162.

Van Hellemond, J.J., Opperdoes, F.R., and Tielens, A.G.M. (2005). The extraordinary mitochondrion and unusual citric acid cycle in Trypanosoma brucei. Biochem. Soc. Trans. *33*, 967–971.

Hird, F.J. (1986). The importance of arginine in evolution. Comp. Biochem. Physiol. B *85*, 285–288.

Hird, F.J.R., Cianciosi, S.C., McLean, R.M., and Niekrash, R.E. (1983). On the possible significance of the transamidination reaction in evolution. Comp. Biochem. Physiol. Part B Comp. Biochem. *76*, 489–495.

Hirumi, H., and Hirumi, K. (1989). Continuous cultivation of Trypanosoma brucei blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. J. Parasitol. *75*, 985–989.

Holzmuller, P., Grébaut, P., Peltier, J.-B., Brizard, J.-P., Perrone, T., Gonzatti, M., Bengaly, Z., Rossignol, M., Aso, P.M., Vincendeau, P., et al. (2008). Secretome of animal trypanosomes. Ann. N. Y. Acad. Sci. *1149*, 337–342.

Horn, D. (2004). The molecular control of antigenic variation in Trypanosoma brucei. Curr. Mol. Med. *4*, 563–576.

Horn, D., and McCulloch, R. (2010). Molecular mechanisms underlying the control of antigenic variation in African trypanosomes. Curr. Opin. Microbiol. *13*, 700–705.

Hsu, J.-L., Huang, S.-Y., Chow, N.-H., and Chen, S.-H. (2003). Stable-isotope dimethyl labeling for quantitative proteomics. Anal. Chem. *75*, 6843–6852.

Hu, Q., Noll, R.J., Li, H., Makarov, A., Hardman, M., and Graham Cooks, R. (2005). The Orbitrap: a new mass spectrometer. J. Mass Spectrom. JMS *40*, 430–443.

Huet, G., Richet, C., Demeyer, D., Bisiau, H., Soudan, B., Tetaert, D., Han, K.K., and Degand, P. (1992). Characterization of different proteolytic activities in Trypanosoma brucei brucei. Biochim. Biophys. Acta *1138*, 213–221.

Jackson, A.P., Sanders, M., Berry, A., McQuillan, J., Aslett, M.A., Quail, M.A., Chukualim, B., Capewell, P., MacLeod, A., Melville, S.E., et al. (2010). The genome sequence of Trypanosoma brucei gambiense, causative agent of chronic human african trypanosomiasis. PLoS Negl. Trop. Dis. *4*, e658.

Jennings, F.W., Rodgers, J., Bradley, B., Gettinby, G., Kennedy, P.G.E., and Murray, M. (2002). Human African trypanosomiasis: Potential therapeutic benefits of an alternative suramin and melarsoprol regimen. Parasitol. Int. *51*, 381–388.

Joe, Y.A., Wolff, E.C., and Park, M.H. (1995). Cloning and expression of human deoxyhypusine synthase cDNA. Structure-function studies with the recombinant enzyme and mutant proteins. J. Biol. Chem. *270*, 22386–22392.

Jones, N.G., Thomas, E.B., Brown, E., Dickens, N.J., Hammarton, T.C., and Mottram, J.C. (2014). Regulators of Trypanosoma brucei cell cycle progression and differentiation identified using a kinome-wide RNAi screen. PLoS Pathog. *10*, e1003886.

Kamleh, A., Barrett, M.P., Wildridge, D., Burchmore, R.J.S., Scheltema, R.A., and Watson, D.G. (2008). Metabolomic profiling using Orbitrap Fourier transform mass spectrometry with hydrophilic interaction chromatography: a method with wide applicability to analysis of biomolecules. Rapid Commun. Mass Spectrom. RCM *22*, 1912–1918.

Kennedy, P.G.E. (2004). Human African trypanosomiasis of the CNS: current issues and challenges. J. Clin. Invest. *113*, 496–504.

Kerkhoven, E. J. (2012). Extending a dynamic mathematical model of metabolism in Trypanosoma brucei. Ph.D. thesis, University of Glasgow, Glasgow.

Keseler, I.M., Bonavides-Martínez, C., Collado-Vides, J., Gama-Castro, S., Gunsalus, R.P., Johnson, D.A., Krummenacker, M., Nolan, L.M., Paley, S., Paulsen, I.T., et al. (2009). EcoCyc: a comprehensive view of Escherichia coli biology. Nucleic Acids Res. *37*, D464–D470.

Kieft, R., Capewell, P., Turner, C.M.R., Veitch, N.J., MacLeod, A., and Hajduk, S. (2010). Mechanism of Trypanosoma brucei gambiense (group 1) resistance to human trypanosome lytic factor. Proc. Natl. Acad. Sci. U. S. A. *107*, 16137–16141.

Kilgour, V. (1980). Trypanosoma: intricacies of biochemistry, morphology and environment. Int. J. Biochem. *12*, 325–332.

Kim, H., Kim, W.-J., Jeon, S.-T., Koh, E.-M., Cha, H.-S., Ahn, K.-S., and Lee, W.-H. (2005). Cyclophilin A may contribute to the inflammatory processes in rheumatoid arthritis through induction of matrix degrading enzymes and inflammatory cytokines from macrophages. Clin. Immunol. Orlando Fla *116*, 217–224.

T' Kindt, R., Jankevics, A., Scheltema, R.A., Zheng, L., Watson, D.G., Dujardin, J.-C., Breitling, R., Coombs, G.H., and Decuypere, S. (2010). Towards an unbiased metabolic profiling of protozoan parasites: optimisation of a Leishmania sampling protocol for HILIC-orbitrap analysis. Anal. Bioanal. Chem. *398*, 2059–2069.

Kingdon, K.H. (1923). A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures. Phys. Rev. *21*, 408–418.

Koek, M.M., Jellema, R.H., van der Greef, J., Tas, A.C., and Hankemeier, T. (2011). Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. Metabolomics Off. J. Metabolomic Soc. *7*, 307–328.

Kohl, L., Robinson, D., and Bastin, P. (2003). Novel roles for the flagellum in cell morphogenesis and cytokinesis of trypanosomes. EMBO J. *22*, 5336–5346.

Lamour, N., Rivière, L., Coustou, V., Coombs, G.H., Barrett, M.P., and Bringaud, F. (2005). Proline metabolism in procyclic Trypanosoma brucei is down-regulated in the presence of glucose. J. Biol. Chem. *280*, 11902–11910.

Langousis, G., and Hill, K.L. (2014). Motility and more: the flagellum of Trypanosoma brucei. Nat. Rev. Microbiol. *12*, 505–518.

Larrouy-Maumus, G., Biswas, T., Hunt, D.M., Kelly, G., Tsodikov, O.V., and de Carvalho, L.P.S. (2013). Discovery of a glycerol 3-phosphate phosphatase reveals glycerophospholipid polar head recycling in Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. U. S. A. *110*, 11320–11325.

Lee, S.H., Stephens, J.L., Paul, K.S., and Englund, P.T. (2006). Fatty acid synthesis by elongases in trypanosomes. Cell *126*, 691–699.

Lee, S.H., Stephens, J.L., and Englund, P.T. (2007). A fatty-acid synthesis mechanism specialized for parasitism. Nat. Rev. Microbiol. *5*, 287–297.

Liesener, A., and Karst, U. (2005). Monitoring enzymatic conversions by mass spectrometry: a critical review. Anal. Bioanal. Chem. *382*, 1451–1464.

Lillico, S., Field, M.C., Blundell, P., Coombs, G.H., and Mottram, J.C. (2003). Essential Roles for GPI-anchored Proteins in African Trypanosomes Revealed Using Mutants Deficient in GPI8. Mol. Biol. Cell *14*, 1182–1194.

Liscombe, D.K., Usera, A.R., and O'Connor, S.E. (2010). Homolog of tocopherol C methyltransferases catalyzes N methylation in anticancer alkaloid biosynthesis. Proc. Natl. Acad. Sci. U. S. A. *107*, 18793–18798.

Mackenzie, N.E., Hall, J.E., Flynn, I.W., and Scott, A.I. (1983). 13C nuclear magnetic resonance studies of anaerobic glycolysis in Trypanosoma brucei spp. Biosci. Rep. *3*, 141–151.

Makarov, A. (2000). Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. Anal. Chem. *72*, 1156–1162.

Makarov, A., Denisov, E., Lange, O., and Horning, S. (2006). Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. J. Am. Soc. Mass Spectrom. *17*, 977–982.

Matthews, K.R. (2005). The developmental cell biology of Trypanosoma brucei. J. Cell Sci. *118*, 283–290.

Mazet, M., Morand, P., Biran, M., Bouyssou, G., Courtois, P., Daulouède, S., Millerioux, Y., Franconi, J.-M., Vincendeau, P., Moreau, P., et al. (2013). Revisiting the central metabolism of the bloodstream forms of Trypanosoma brucei: production of acetate in the mitochondrion is essential for parasite viability. PLoS Negl. Trop. Dis. *7*, e2587.

Melville, S.E., Leech, V., Gerrard, C.S., Tait, A., and Blackwell, J.M. (1998). The molecular karyotype of the megabase chromosomes of Trypanosoma brucei and the assignment of chromosome markers. Mol. Biochem. Parasitol. *94*, 155–173.

Melville, S.E., Leech, V., Navarro, M., and Cross, G.A. (2000). The molecular karyotype of the megabase chromosomes of Trypanosoma brucei stock 427. Mol. Biochem. Parasitol. *111*, 261–273.

Michels, P.A.M., Bringaud, F., Herman, M., and Hannaert, V. (2006). Metabolic functions of glycosomes in trypanosomatids. Biochim. Biophys. Acta *1763*, 1463–1477.

Miézan, T.W., Bronner, U., Doua, F., Cattand, P., and Rombo, L. (1994). Long-term exposure of Trypanosoma brucei gambiense to pentamidine in vitro. Trans. R. Soc. Trop. Med. Hyg. *88*, 332–333.

Miranda, M.R., Canepa, G.E., Bouvier, L.A., and Pereira, C.A. (2006). Trypanosoma cruzi: Oxidative stress induces arginine kinase expression. Exp. Parasitol. *114*, 341–344.

Molina-Portela, M. del P., Lugli, E.B., Recio-Pinto, E., and Raper, J. (2005). Trypanosome lytic factor, a subclass of high-density lipoprotein, forms cation-selective pores in membranes. Mol. Biochem. Parasitol. *144*, 218–226.

Moreno, B., Urbina, J.A., Oldfield, E., Bailey, B.N., Rodrigues, C.O., and Docampo, R. (2000). 31P NMR Spectroscopy of Trypanosoma brucei, Trypanosoma cruzi, and Leishmania major EVIDENCE FOR HIGH LEVELS OF CONDENSED INORGANIC PHOSPHATES. J. Biol. Chem. *275*, 28356–28362.

Morris, S.M. (2002). Regulation of Enzymes of the Urea Cycle and Arginine Metabolism. Annu. Rev. Nutr. *22*, 87–105.

Morris, M.T., DeBruin, C., Yang, Z., Chambers, J.W., Smith, K.S., and Morris, J.C. (2006). Activity of a second Trypanosoma brucei hexokinase is controlled by an 18-amino-acid C-terminal tail. Eukaryot. Cell *5*, 2014–2023.

Morty, R.E., Lonsdale-Eccles, J.D., Mentele, R., Auerswald, E.A., and Coetzer, T.H. (2001). Trypanosome-derived oligopeptidase B is released into the plasma of infected rodents, where it persists and retains full catalytic activity. Infect. Immun. *69*, 2757–2761.

Nerima, B., Nilsson, D., and Mäser, P. (2010). Comparative genomics of metabolic networks of free-living and parasitic eukaryotes. BMC Genomics *11*, 217.

Nguyen, S., Jones, D.C., Wyllie, S., Fairlamb, A.H., and Phillips, M.A. (2013). Allosteric activation of trypanosomatid deoxyhypusine synthase by a catalytically dead paralog. J. Biol. Chem. *288*, 15256–15267.

Nicholson, J.K., Connelly, J., Lindon, J.C., and Holmes, E. (2002). Metabonomics: a platform for studying drug toxicity and gene function. Nat. Rev. Drug Discov. *1*, 153–161.

Nikolskaia, O.V., de A Lima, A.P.C., Kim, Y.V., Lonsdale-Eccles, J.D., Fukuma, T., Scharfstein, J., and Grab, D.J. (2006). Blood-brain barrier traversal by African trypanosomes requires calcium signaling induced by parasite cysteine protease. J. Clin. Invest. *116*, 2739–2747.

Nozaki, T., Ali, V., and Tokoro, M. (2005). Sulfur-containing amino acid metabolism in parasitic protozoa. Adv. Parasitol. *60*, 1–99.

Nwagwu, M., and Opperdoes, F.R. (1982). Regulation of glycolysis in Trypanosoma brucei: hexokinase and phosphofructokinase activity. Acta Trop. *39*, 61–72.

O'Farrell, P.H. (1975). High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. *250*, 4007–4021.

Olver, C., and Vidal, M. (2007). Proteomic analysis of secreted exosomes. Subcell. Biochem. *43*, 99–131.

Opperdoes, F.R., and Borst, P. (1977). Localization of nine glycolytic enzymes in a microbody-like organelle in Trypanosoma brucei: the glycosome. FEBS Lett. *80*, 360–364.

Ozment, C., Barchue, J., DeLucas, L.J., and Chattopadhyay, D. (1999). Structural Study of Escherichia coli NAD Synthetase: Overexpression, Purification, Crystallization, and Preliminary Crystallographic Analysis. J. Struct. Biol. *127*, 279–282.

Parkin, D.W. (1996). Purine-specific nucleoside N-ribohydrolase from Trypanosoma brucei brucei. Purification, specificity, and kinetic mechanism. J. Biol. Chem. *271*, 21713–21719.

Peacock, L., Ferris, V., Bailey, M., and Gibson, W. (2008). Fly transmission and mating of Trypanosoma brucei brucei strain 427. Mol. Biochem. Parasitol. *160*, 100–106.

Pegg, A.E. (2009). Mammalian polyamine metabolism and function. IUBMB Life *61*, 880–894.

Pereira, C.A. (2014). Arginine kinase: a potential pharmacological target in trypanosomiasis. Infect. Disord. Drug Targets *14*, 30–36.

Pils, B., and Schultz, J. (2004). Inactive enzyme-homologues find new function in regulatory processes. J. Mol. Biol. *340*, 399–404.

Poulin, R., Lu, L., Ackermann, B., Bey, P., and Pegg, A.E. (1992). Mechanism of the irreversible inactivation of mouse ornithine decarboxylase by alpha-difluoromethylornithine. Characterization of sequences at the inhibitor and coenzyme binding sites. J. Biol. Chem. *267*, 150–158.

Priotto, G., Kasparian, S., Ngouama, D., Ghorashian, S., Arnold, U., Ghabri, S., and Karunakara, U. (2007). Nifurtimox-Eflornithine Combination Therapy for Second-Stage Trypanosoma brucei gambiense Sleeping Sickness: A Randomized Clinical Trial in Congo. Clin. Infect. Dis. *45*, 1435–1442.

Probst, P., Stromberg, E., Ghalib, H.W., Mozel, M., Badaro, R., Reed, S.G., and Webb, J.R. (2001). Identification and Characterization of T Cell-Stimulating Antigens from Leishmania by CD4 T Cell Expression Cloning. J. Immunol. *166*, 498–505.

Prosser, G.A., Larrouy-Maumus, G., and de Carvalho, L.P.S. (2014). Metabolomic strategies for the identification of new enzyme functions and metabolic pathways. EMBO Rep. *15*, 657–669.

Proto, W.R., Castanys-Munoz, E., Black, A., Tetley, L., Moss, C.X., Juliano, L., Coombs, G.H., and Mottram, J.C. (2011). Trypanosoma brucei metacaspase 4 is a pseudopeptidase and a virulence factor. J. Biol. Chem. *286*, 39914–39925.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., et al. (2012). The Pfam protein families database. Nucleic Acids Res. *40*, D290–D301.

Quiñones, W., Peña, P., Domingo-Sananes, M., Cáceres, A., Michels, P.A.M., Avilan, L., and Concepción, J.L. (2007). Leishmania mexicana: Molecular cloning and characterization of enolase. Exp. Parasitol. *116*, 241–251.

Ranganathan, S., and Garg, G. (2009). Secretome: clues into pathogen infection and clinical applications. Genome Med. *1*, 113.

Redmond, S., Vadivelu, J., and Field, M.C. (2003). RNAit: an automated web-based tool for the selection of RNAi targets in Trypanosoma brucei. Mol. Biochem. Parasitol. *128*, 115–118.

Reguera, R.M., Redondo, C.M., Pérez-Pertejo, Y., and Balaña-Fouce, R. (2007). S-Adenosylmethionine in protozoan parasites: Functions, synthesis and regulation. Mol. Biochem. Parasitol. *152*, 1–10.

Rivière, L., Weelden, S.W.H. van, Glass, P., Vegh, P., Coustou, V., Biran, M., Hellemond, J.J. van, Bringaud, F., Tielens, A.G.M., and Boshart, M. (2004). Acetyl:Succinate CoA-transferase in Procyclic Trypanosoma brucei GENE IDENTIFICATION AND ROLE IN CARBOHYDRATE METABOLISM. J. Biol. Chem. *279*, 45337–45346.

Robinson, M.D., Souza, D.P.D., Keen, W.W., Saunders, E.C., McConville, M.J., Speed, T.P., and Likić, V.A. (2007). A dynamic programming approach for the alignment of signal peaks in multiple gas chromatography-mass spectrometry experiments. BMC Bioinformatics *8*, 419.

Rodgers, J. (2009). Human African trypanosomiasis, chemotherapy and CNS disease. J. Neuroimmunol. *211*, 16–22.

Saghatelian, A., Trauger, S.A., Want, E.J., Hawkins, E.G., Siuzdak, G., and Cravatt, B.F. (2004). Assignment of endogenous substrates to enzymes by global metabolite profiling. Biochemistry (Mosc.) *43*, 14332–14339.

Saito, N., Robert, M., Kitamura, S., Baran, R., Soga, T., Mori, H., Nishioka, T., and Tomita, M. (2006). Metabolomics approach for enzyme discovery. J. Proteome Res. *5*, 1979–1987.

Saito, N., Robert, M., Kochi, H., Matsuo, G., Kakazu, Y., Soga, T., and Tomita, M. (2009). Metabolite profiling reveals YihU as a novel hydroxybutyrate dehydrogenase for alternative succinic semialdehyde metabolism in Escherichia coli. J. Biol. Chem. *284*, 16442–16451.

Schauer, N., Steinhauser, D., Strelkov, S., Schomburg, D., Allison, G., Moritz, T., Lundgren, K., Roessner-Tunali, U., Forbes, M.G., Willmitzer, L., et al. (2005). GC–MS libraries for the rapid identification of metabolites in complex biological samples. FEBS Lett. *579*, 1332–1337.

Scheltema, R.A., Jankevics, A., Jansen, R.C., Swertz, M.A., and Breitling, R. (2011). PeakML/mzMatch: a file format, Java library, R library, and tool-chain for mass spectrometry data analysis. Anal. Chem. *83*, 2786–2793.

Schumann Burkard, G., Jutzi, P., and Roditi, I. (2011). Genome-wide RNAi screens in bloodstream form trypanosomes identify drug transporters. Mol. Biochem. Parasitol. *175*, 91–94.

Sherwin, T., and Gull, K. (1989). Visualization of detyrosination along single microtubules reveals novel mechanisms of assembly during cytoskeletal duplication in trypanosomes. Cell *57*, 211–221.

Da Silveira, J.F., Abrahamsohn, P.A., and Colli, W. (1979). Plasma membrane vesicles isolated from epimastigote forms of Trypanosoma cruzi. Biochim. Biophys. Acta *550*, 222–232.

Silverman, J.S., and Bangs, J.D. (2012). Form and function in the trypanosomal secretory pathway. Curr. Opin. Microbiol. *15*, 463–468.

Simarro, P.P., Cecchi, G., Franco, J.R., Paone, M., Diarra, A., Ruiz-Postigo, J.A., Fèvre, E.M., Mattioli, R.C., and Jannin, J.G. (2012). Estimating and mapping the population at risk of sleeping sickness. PLoS Negl. Trop. Dis. *6*, e1859.

Simpson, L., and Shaw, J. (1989). RNA editing and the mitochondrial cryptogenes of kinetoplastid protozoa. Cell *57*, 355–366.

Smith, T.K., and Bütikofer, P. (2010). Lipid metabolism in Trypanosoma brucei. Mol. Biochem. Parasitol. *172*, 66–79.

Spencer, R.L., and Preiss, J. (1967). Biosynthesis of diphosphopyridine nucleotide. The purification and the properties of diphospyridine nucleotide synthetase from Escherichia coli b. J. Biol. Chem. *242*, 385–392.

Stephens, J.L., Lee, S.H., Paul, K.S., and Englund, P.T. (2007). Mitochondrial fatty acid synthesis in Trypanosoma brucei. J. Biol. Chem. *282*, 4427–4436.

Stipanuk, M.H. (2004). SULFUR AMINO ACID METABOLISM: Pathways for Production and Removal of Homocysteine and Cysteine. Annu. Rev. Nutr. *24*, 539–577.

Stuart, K.D., and Gelvin, S.B. (1982). Localization of kinetoplast DNA maxicircle transcripts in bloodstream and procyclic form Trypanosoma brucei. Mol. Cell. Biol. *2*, 845–852.

Sutton, R.E., and Boothroyd, J.C. (1986). Evidence for Trans splicing in trypanosomes. Cell *47*, 527–535.

Tabor, C.W., and Tabor, H. (1984). Polyamines. Annu. Rev. Biochem. *53*, 749–790.

Tetaud, E., Lecuix, I., Sheldrake, T., Baltz, T., and Fairlamb, A.H. (2002). A new expression vector for Crithidia fasciculata and Leishmania. Mol. Biochem. Parasitol. *120*, 195–204.

Tjalsma, H., Bolhuis, A., Jongbloed, J.D., Bron, S., and van Dijl, J.M. (2000). Signal peptide-dependent protein transport in Bacillus subtilis: a genome-based survey of the secretome. Microbiol. Mol. Biol. Rev. MMBR *64*, 515–547.

Tomlinson, S., Jansen, A.M., Koudinov, A., Ghiso, J.A., Choi-Miura, N.H., Rifkin, M.R., Ohtaki, S., and Nussenzweig, V. (1995). High-density-lipoprotein-independent killing of Trypanosoma brucei by human serum. Mol. Biochem. Parasitol. *70*, 131–138.

Triggs, V.P., and Bangs, J.D. (2003). Glycosylphosphatidylinositol-Dependent Protein Trafficking in Bloodstream Stage Trypanosoma brucei. Eukaryot. Cell *2*, 76–83.

Vanegas, G., Quiñones, W., Carrasco-López, C., Concepción, J.L., Albericio, F., and Avilán, L. (2007). Enolase as a plasminogen binding protein in Leishmania mexicana. Parasitol. Res. *101*, 1511–1516.

Vanhamme, L., Paturiaux-Hanocq, F., Poelvoorde, P., Nolan, D.P., Lins, L., Van Den Abbeele, J., Pays, A., Tebabi, P., Van Xong, H., Jacquet, A., et al. (2003). Apolipoprotein L-I is the trypanosome lytic factor of human serum. Nature *422*, 83–87.

Velez, N., Brautigam, C.A., and Phillips, M.A. (2013). Trypanosoma brucei S-adenosylmethionine decarboxylase N terminus is essential for allosteric activation by the regulatory subunit prozyme. J. Biol. Chem. *288*, 5232–5240.

Vincent, I. M. (2011). Using metabolomic analyses to study mode of action and resistance to Eflornithine in Trypanosoma brucei. Ph.D. thesis, University of Glasgow, Glasgow.

Vincent, I.M., Creek, D.J., Burgess, K., Woods, D.J., Burchmore, R.J.S., and Barrett, M.P. (2012). Untargeted metabolomics reveals a lack of synergy between nifurtimox and eflornithine against Trypanosoma brucei. PLoS Negl. Trop. Dis. *6*, e1618.

Voncken, F., Gao, F., Wadforth, C., Harley, M., and Colasante, C. (2013). The phosphoarginine energy-buffering system of trypanosoma brucei involves multiple arginine kinase isoforms with different subcellular locations. PloS One *8*, e65908.

Walker, J., and Barrett, J. (1997). Parasite sulphur amino acid metabolism. Int. J. Parasitol. *27*, 883–897.

Watson, D.G. (2010). The potential of mass spectrometry for the global profiling of parasite metabolomes. Parasitology *137*, 1409–1423.

Van Weelden, S.W.H., van Hellemond, J.J., Opperdoes, F.R., and Tielens, A.G.M. (2005). New functions for parts of the Krebs cycle in procyclic Trypanosoma brucei, a cycle not operating as a cycle. J. Biol. Chem. *280*, 12451–12460.

Van der Werf, M.J., Jellema, R.H., and Hankemeier, T. (2005). Microbial metabolomics: replacing trial-and-error by the unbiased selection and ranking of targets. J. Ind. Microbiol. Biotechnol. *32*, 234–252.

Whisstock, J.C., and Lesk, A.M. (2003). Prediction of protein function from protein sequence and structure. Q. Rev. Biophys. *36*, 307–340.

Wickstead, B., Ersfeld, K., and Gull, K. (2004). The small chromosomes of Trypanosoma brucei involved in antigenic variation are constructed around repetitive palindromes. Genome Res. *14*, 1014–1024.

Willert, E., and Phillips, M.A. (2012). Regulation and function of polyamines in African trypanosomes. Trends Parasitol. *28*, 66–72.

Willert, E.K., Fitzpatrick, R., and Phillips, M.A. (2007). Allosteric regulation of an essential trypanosome polyamine biosynthetic enzyme by a catalytically dead homolog. Proc. Natl. Acad. Sci. U. S. A. *104*, 8275–8280.

Wilson, W., and Seebeck, T. (1997). Identification of a profilin homologue in Trypanosoma brucei by complementation screening. Gene *187*, 201–209.

Winder, C.L., Dunn, W.B., and Goodacre, R. (2011). TARDIS-based microbial metabolomics: time and relative differences in systems. Trends Microbiol. *19*, 315–322.

Wojcik, M., Seidle, H.F., Bieganowski, P., and Brenner, C. (2006). Glutamine-dependent NAD+ synthetase. How a two-domain, three-substrate enzyme avoids waste. J. Biol. Chem. *281*, 33395–33402.

Appendix A

Growth media for *T. b.brucei*

(1) Modified HMI-9

For 500 ml modified HMI-9:

Iscoves modified Dulbecco's medium + glutamax (Gibco)	365 ml
Methyl cellulose (Sigma)	0.55g

add:

Compound	**Concentration (mM) Stocksolution**	**Concentration final in modified HMI-9**
Bathocuproinedisulfonic acid	5	0.5 µM
L-cysteine	100	1 mM
Sodium pyruvate	200	2 M
Thymidine	16	160 µM
Hypoxanthine		2 mM
D-glucose		5.5 mM
Adenosine		0.5 mM
Guanosine		0.5 mM
Mercaptoethanol		3 µM
Pen/Strep		5000 Units
FBS Gold (PAA)		20%
Serum plus		20%

Table A1: Modified HMI-9 (adapted from Paul Voorheis).

(2) Creek's minimal medium (CMM)

Compound	Concentration (µM)
Bathocuproine disulfonic acid	52
Phenolsulfonphtalein	42
HEPES	10,000
NaCl	77,590
$CaCl_2$	1,490
KCl	4,400
$MgSO_4$	814
$NaHCO_3$	35,950
D-glucose	10,000
L-glutamine	1,000
L-cysteine	1,000
Mercaptoethanol	192
FBS Gold (PAA)	10%

Table A2: Creek's minimal media (CMM) (Creel et al., 2013). If a different brand FBS does not support growth in this medium, L-arginine, L-tyrosine, L-methionine, L-leucine, L-phenylalanine and L-tryptophan should be added (100µM).

Appendix B

Gene ID	Gene name	Description	Sequence
Tb427.01.1130	glycerol-3-phosphate dehydrogenase	Recombinat expression, fwd	GGTATTGAGGGTCGCATGGGTCGCTATACGCGCCG
		Recombinat Expresssion, rev	AGAGGAGAGTTAGAGCCTCATGTGAGAGCCGTGGCGG
Tb427.06.4920	S-adenosylmethionine synthetase	Recombinat expression, fwd	GGTATTGAGGGTCGCATGTCCGTGCGCCAG
		Recombinat Expresssion, rev	AGAGGAGAGTTAGAGCCCTACTGCACGTCACTAAGACC
Tb427.07.4390	threonine synthase	Recombinat expression, fwd	GGTATTGAGGGTCGCATGCTCACCTTACGCG
		Recombinat Expresssion, rev	AGAGGAGAGTTAGAGCCTCAGCAAGAGCCCTCACTGG
Tb427.08.3800	nucleoside phosphatase	Recombinat expression, fwd	GGTATTGAGGGTCGCATGTTGCACTCAAACCACCC
		Recombinat Expresssion, rev	AGAGGAGAGTTAGAGCCTCACACACAGTGCTCCTCTAAGC
Tb427.10.12980	methyltransferase	Recombinat expression, fwd	GGTATTGAGGGTCGCATGTGGAAACCGAAACAGCG
		Recombinat Expresssion, rev	AGAGGAGAGTTAGAGCCTCAAGGATCACGCACGTCTGGC
Tb427.10.13130	UTP-glucose-1-phosphate uridylyltransferase	Recombinat expression, fwd	GGTATTGAGGGTCGCATGCCGCTAAACCCTCCTTCAGC
		Recombinat Expresssion, rev	AGAGGAGAGTTAGAGCCCTACTCGACTACCACAACC
Tb427.10.13430	citrate synthase	Recombinat expression, fwd	GGTATTGAGGGTCGCATGTGCATGCGTGCTCG
		Recombinat Expresssion, rev	AGAGGAGAGTTAGAGCCCTACGCTATGTTGTACTTTGTG
Tb427tmp.01.3640	acyl-CoA dehydrogenase	Recombinat expression, fwd	GGTATTGAGGGTCGCATGTTTCGTCGTAGCCTTTCCC
		Recombinat Expresssion, rev	AGAGGAGAGTTAGAGCCTCACCTAAGATCCCATGTCG
Tb427tmp.02.0530	phosphoribosylpyrophosphate synthetase	Recombinat expression, fwd	GGTATTGAGGGTCGCATGGGTTGTGCCATGCATTTCGC
		Recombinat Expresssion, rev	AGAGGAGAGTTAGAGCCCTATGTGCGGAACAAAGAGG
Tb427tmp.02.3040	aldo/keto reductase	Recombinat expression, fwd	GGTATTGAGGGTCGCATGGACCGTATTCCATATTTGG
		Recombinat Expresssion, rev	AGAGGAGAGTTAGAGCCTTAATCTATCGTGTTGCTATGCC
Tb427.10.2490	glucose-6-phosphate 1-dehydrogenase	Recombinat expression, fwd	GGTATTGAGGGTCGCATGGACGGTGATCTTTCCC
		Recombinat Expresssion, rev	AGAGGAGAGTTAGAGCCTTACAAATGATGAAGCTTCCGC

Gene ID	Gene name	Description	Sequence
Tb427.05.3820	aspartate carbamoyltransferase	Recombinat expression, fwd	GGTATTGAGGGTCGCATGGCGGAGCTGCAACCTG
		Recombinat Expresssion, rev	AGAGGAGAGTTAGAGCCTTAGGCGAGAACACTATAAAG
Tb427.10.2010	Hexokinase I	Recombinat expression, fwd	GGTATTGAGGGTCGCATGTCTAGACGCCTAAACAATATCC
		Recombinat Expresssion, rev	AGAGGAGAGTTAGAGCCTTACTTGTCGTTCACCACCATTGCG
Tb427.05.4560	guanine deaminase	Recombinat expression, fwd	GGTATTGAGGGTCGCATGNAAACCGCCGTGCCTGTGCG
		Recombinat Expresssion, rev	AGAGGAGAGTTAGAGCCCTAAAGTTCTACCCTCCCGTGGACC

Table B1: Oligonucleotides used in this study for protein over expression. The inserted overhang necessary for ligase independent cloning is underlined.

Gene ID	Gene name	Description	Sequence
Tb427.05.3820	aspartate carbamoyltransferase	G197^{RNAi}, fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTGTGACGTACTTGTGCTGCGT
		G197^{RNAi}, rev	GGGGACCACTTTGTACAAGAAAGCTGGGTAATGACGATATCCGCCTTTG
Tb427tmp.02.3040	aldo/keto reductase	G196^{RNAi}, fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTCATTGATTACGCCGATTGTG
		G196^{RNAi}, rev	GGGGACCACTTTGTACAAGAAAGCTGGGTGCTCCCTCCGAGAAGAAACT
Tb11.01.6500	NAD+ synthase	G162^{RNAi}, fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGGTTGGAGGAATACCGAAG
		G162^{RNAi}, rev	GGGGACCACTTTGTACAAGAAAGCTGGGTGTAGCCATCCTCGTCCATGT
Tb927.7.5680	deoxyribose-phosphate aldolase	G159^{RNAi}, fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTCTAAAACCTGAGGCAACCCA
		G159^{RNAi}, rev	GGGGACCACTTTGTACAAGAAAGCTGGGTGCAGCGACTGACAGGATACA

Table B2: Oligonucleotides used in this study for creation of RNAi lines

Appendix C

General buffers and solutions

0.1% coomassie blue stain (200 mL)

Coomassie blue	0.2 g
Methanol	80 mL
Acetic acid	20 mL
dH_2O	100 mL

Destaining solution for protein gels (500 mL)

Methanol	100 mL
Acetic acid	50 mL
dH_2O	350 mL

LB medium (Luria Bertani broth, pH 7)

LB powder (Sigma-Aldrich)	25 g
dH2O	1 L

LB agar

Luria Agar (Sigma-Aldrich)	35 g
dH_2O	1 L

DiGE lysis buffer:

Urea	6 M
Thiourea	2 M
CHAPS	4%
Tris base	25mM

Rehydration buffer:

Urea	6 M
Thiourea	2 M
CHAPS	4%
IPG buffer	0.5%
DTT	65 mM
Trace of bromophenol blue	

Appendix D

Table of the three authentic standard mixes run with every metabolomics experiment. Shown are the metabolite name and formula. The polarity mode (positive or negative) the metabolite was best detected in and the expected retention time.

Mix	Compound Name	Formula	Polarity	Expected RT
1	1-Naphthylacetic acid	C12H10O2	+	4.21
	Serotonin	C10H12N2O	+	4.28
	Melatonin	C13H16N2O2	+	4.67
	Phenylhydrazine	C6H8N2	+	5.08
	Nicotinate	C6H5NO2	+	6.99
	Pyridoxine	C8H11NO3	+	7.45
	riboflavin	C17H20N4O6	+	7.62
	Glycerol	C3H8O3	+	8.12
	Adenine	C5H5N5	+	8.34
	Creatinine	C4H7N3O	+	8.42
	L-Phenylalanine	C9H11NO2	+	8.34
	L-Leucine	C6H13NO2	+	8.74
	4-Aminobenzoate	C7H7NO2	+	9.35
	Inosine	C10H12N4O5	+	8.96
	2-Phenylglycine	C8H9NO2	+	8.9
	Selenomethionine	C5H11NO2Se	+	9.03
	L-Methionine	C5H11NO2S	+	9.07
	Imidazole-4-acetate	C5H6N2O2	+	9.03
	L-Tryptophan	C11H12N2O2	+	9.35
	N-Acetyl-D-glucosamine	C8H15NO6	+	9.33
	Cytidine	C9H13N3O5	+	9.53
	Guanine	C5H5N5O	+	9.98
	L-Valine	C5H11NO2	+	9.59
	L-Proline	C5H9NO2	+	9.75
	sn-glycero-3-Phosphocholine	C8H20NO6P	+	10.23
	L-Threonine	C4H9NO3	+	10.4
	trans-4-Hydroxy-L-proline	C5H9NO3	+	10.52
	L-Histidine	C6H9N3O2	+	10.48
	N2-Acetyl-L-lysine	C8H16N2O3	+	10.57
	D-Glucosamine	C6H13NO5	+	10.72
	L-Glutamine	C5H10N2O3	+	10.65
	Taurine	C2H7NO3S	+	10.94
	L-Aspartate	C4H7NO4	+	10.52
	L-Asparagine	C4H8N2O3	+	10.76
	beta-Alanine	C3H7NO2	+	10.94
	Glycine	C2H5NO2	+	11.13
	dGMP	C10H14N5O7P	+	10.52
	L-Serine	C3H7NO3	+	11.09
	Ethanolamine phosphate	C2H8NO4P	+	10.72

Mix	Compound Name	Formula	Polarity	Expected RT
1	L-Citrulline	C6H13N3O3	+	11.04
	L-Cystine	C6H12N2O4S2	+	10.87
	meso-2_6-Diaminoheptanedioate	C7H14N2O4	+	11.8
	Putrescine	C4H12N2	+	11.56
	L-2_4-Diaminobutanoate	C4H10N2O2	+	12.27
	Thiamin	C12H16N4OS	+	16.71
	L-Lysine	C6H14N2O2	+	15.38
	L-Arginine	C6H14N4O2	+	16.42
	S-Adenosyl-L-methionine	C15H22N6O5S	+	0
	Phenolsulfonphthalein	C19H14O5S	-	5.11
	Thymidine	C10H14N2O5	-	6.79
	MOPS	C7H15NO4S	-	6.73
	Pyruvate	C3H4O3	-	7.41
	Deoxyuridine	C9H12N2O5	-	7.52
	4-Coumarate	C9H8O3	-	7.39
	pyrazinoate	C5H4N2O2	-	7.72
	(R)-3-Hydroxybutanoate	C4H8O3	-	7.87
	3'_5'-Cyclic AMP	C10H12N5O6P	-	7.93
	Uridine	C9H12N2O6	-	8.47
	D-glucose	C6H12O6	-	10.77
	Orotate	C5H4N2O4	-	8.58
	L-Rhamnose	C6H12O5	-	9.16
	Xanthine	C5H4N4O2	-	9.28
	3_4-Dihydroxyphenylacetate	C8H8O4	-	10.15
	Pyridoxal phosphate	C8H10NO6P	-	0
	Phthalate	C8H6O4	-	10.24
	D-Gluconic acid	C6H12O7	-	10.17
	UMP	C9H13N2O9P	-	9.93
	Thiamin diphosphate	C12H18N4O7P2S	-	9.86
	L-Glutamate	C5H9NO4	-	10.34
	IMP	C10H13N4O8P	-	10.15
	Methylmalonate	C4H6O4	-	10.54
	CMP	C9H14N3O8P	-	10.43
	(R)-2-Hydroxyglutarate	C5H8O5	-	10.62
	L-Cysteate	C3H7NO5S	-	11.08
	D-ribose 5-phosphate	C5H11O8P	-	10.66
	Malonate	C3H4O4	-	10.95
	D-Galacturonate	C6H10O7	-	11.28
	Fumarate	C4H4O4	-	9.89
	D-glucose 6-phosphate	C6H13O9P	-	11.12
	D-Galactarate	C6H10O8	-	11.48
	Phosphoenolpyruvate	C3H5O6P	-	11.52
	6-Phospho-D-gluconate	C6H13O10P	-	11.39
	cis-Aconitate	C6H6O6	-	11.77
	Gallate	C7H6O5	-	13.11
	D-Erythrose	C4H8O4	-	0
	Maltose	C12H22O11	-	0

Mix	Compound Name	Formula	Polarity	Expected RT
	2-oxobutanoate	C4H6O3	-	0
1	2-Methylcitrate	C7H10O7	-	0
	Glyoxylate	C2H2O3	-	0
	glycolate	C2H4O3	-	0
	methylglyoxal	C3H4O2	-	0
	IDP	C10H14N4O11P2	-	0
	ADP	C10H15N5O10P2	-	0
	D-Fructose 1_6-bisphosphate	C6H14O12P2	-	0
	CoA	C21H36N7O16P3S	-	0
	2_3-Bisphospho-D-glycerate	C3H8O10P2	-	0
	3-Hydroxyphenylacetate	C8H8O3	-	0
	Cholesterol	C27H46O	+	0
	1-Butanol	C4H10O	+	0
	Oxaloacetate	C4H4O5	+	0
	Imidazole	C3H4N2	+	0
	glyceraldehyde	C3H6O3	+	0
2	Menadione	C11H8O2	+	15.65
	Dopamine	C8H11NO2	+	8.28
	Thiopurine S-methylether	C6H6N4S	+	6.35
	5'-Methylthioadenosine	C11H15N5O3S	+	6.52
	Nicotinamide	C6H6N2O	+	6.8
	4-(beta-Acetylaminoethyl)imidazole	C7H11N3O	+	6.92
	Pyridoxal	C8H9NO3	+	7.74
	6-Methylaminopurine	C6H7N5	+	7.36
	Deoxyadenosine	C10H13N5O3	+	7.46
	Pantothenate	C9H17NO5	+	7.46
	Uracil	C4H4N2O2	+	7.85
	Picolinic acid	C6H5NO2	+	7.55
	thymine	C5H6N2O2	+	7.11
	Adenosine	C10H13N5O4	+	7.94
	Hypoxanthine	C5H4N4O	+	8.75
	5-Oxoproline	C5H7NO3	+	8.36
	HEPES	C8H18N2O4S	+	8.36
	L-Kynurenine	C10H12N2O3	+	8.81
	Betaine	C5H11NO2	+	9.03
	cytosine	C4H5N3O	+	9.39
	FAD	C27H33N9O15P2	+	8.47
	L-isoleucine	C6H13NO2	+	9.3
	Methylcysteine	C4H9NO2S	+	9.18
	Ala-Gly	C5H10N2O3	+	9.14
	Guanosine	C10H13N5O5	+	9.98
	dAMP	C10H14N5O6P	+	9.14
	1-Aminocyclopropane-1-carboxylate	C4H7NO2	+	9.59
	N(pi)-Methyl-L-histidine	C7H11N3O2	+	9.47
	L-Carnitine	C7H15NO3	+	9.89
	Eflornithine	C6H12F2N2O2	+	9.98
	L-Tyrosine	C9H11NO3	+	9.98

Mix	Compound Name	Formula	Polarity	Expected RT
2	4-Trimethylammoniobutanoate	C7H15NO2	+	9.77
	S-Adenosyl-L-homocysteine	C14H20N6O5S	+	9.68
	AMP	C10H14N5O7P	+	9.59
	NAD+	C21H27N7O14P2	+	0
	Glycylglycine	C4H8N2O3	+	9.98
	Phenol	C6H6O	+	0
	D-Glucosamine-6-phosphate	C6H14NO8P	+	10.79
	L-Alanine	C3H7NO2	+	10.73
	Choline phosphate	C5H14NO4P	+	10.21
	dCMP	C9H14N3O7P	+	10.1
	L-2-Aminoadipate	C6H11NO4	+	10.38
	L-homoserine	C4H9NO3	+	10.75
	N-Acetylornithine	C7H14N2O3	+	10.7
	4-Aminobutanoate	C4H9NO2	+	10.94
	L-Cystathionine	C7H14N2O4S	+	11.22
	agmatine	C5H14N4	+	11.15
	Spermine	C10H26N4	+	12.23
	Trypanothione disulfide	C27H47N9O10S2	+	12.17
	Choline	C5H13NO	+	15.65
	1-Aminopropan-2-ol	C3H9NO	+	15.65
	L-Ornithine	C5H12N2O2	+	14.24
	5-Methoxytryptamine	C11H14N2O	+	4.59
	Methylguanidine	C2H7N3	+	0
	Biotin	C10H16N2O3S	+	0
	L-2_3-Diaminopropanoate	C3H8N2O2	+	11.31
	1_3-Diaminopropane	C3H10N2	+	14.04
	Taurocholate	C26H45NO7S	-	4.54
	Lipoate	C8H14O2S2	-	4.54
	Phenylacetylglycine	C10H11NO3	-	5.89
	Phenylpyruvate	C9H8O3	-	4.6
	D-Threose	C4H8O4	-	8.27
	succinate semialdehyde	C4H6O3	-	8.27
	N-Acetylglutamine	C7H12N2O4	-	8.46
	3-(4-Hydroxyphenyl)pyruvate	C9H8O4	-	6.36
	D-Ribose	C5H10O5	-	8.91
	FMN	C17H21N4O9P	-	8.52
	L-Gulono-1_4-lactone	C6H10O6	-	9.31
	Orotidine	C10H12N2O8	-	9.5
	N-Acetylneuraminate	C11H19NO9	-	9.74
	dUMP	C9H13N2O8P	-	9.72
	allantoin	C4H6N4O3	-	11.02
	dIMP	C10H13N4O7P	-	9.95
	sn-Glycerol 3-phosphate	C3H9O6P	-	10.32
	sucrose	C12H22O11	-	10.78
	D-Erythrose 4-phosphate	C4H9O7P	-	10.85
	Succinate	C4H6O4	-	10.61
	D-Glucuronolactone	C6H8O6	-	10.85

Mix	Compound Name	Formula	Polarity	Expected RT
2	Mesaconate	C5H6O4	-	10.8
	2-Oxoglutarate	C5H6O5	-	10.91
	GMP	C10H14N5O8P	-	10.67
	(S)-Malate	C4H6O5	-	11.04
	UDP-glucose	C15H24N2O17P2	-	10.95
	Folate	C19H19N7O6	-	11.1
	DL-Glyceraldehyde 3-phosphate	C3H7O6P	-	0
	NADP+	C21H28N7O17P3	-	10.78
	ATP	C10H16N5O13P3	-	10.85
	3-Phospho-D-glycerate	C3H7O7P	-	11.19
	Oxalate	C2H2O4	-	12.02
	GDP	C10H15N5O11P2	-	11.56
	citrate	C6H8O7	-	11.85
	CTP	C9H16N3O14P3	-	11.67
	GTP	C10H16N5O14P3	-	12.11
	2-Deoxy-D-glucose	C6H12O5	-	8.65
	(R)-Lactate	C3H6O3	-	5.37
	Acetyl-CoA	C23H38N7O17P3S	-	0
3	Dihydrolipoamide	C8H17NOS2	+	8.11
	lipoamide	C8H15NOS2	+	4.94
	2-Aminobutan-4-olide	C4H7NO2	+	6.62
	2-hydroxyethyldisulfide	C4H10O2S2	+	6.8
	Isonicotinic acid	C6H5NO2	+	7.01
	L-cysteine	C3H7NO2S	+	7.16
	L-Noradrenaline	C8H11NO3	+	7.42
	Biopterin	C9H11N5O3	+	7.86
	O-Acetylcarnitine	C9H17NO4	+	8.74
	O-Acetyl-L-serine	C5H9NO4	+	8.55
	L-Homocysteine	C4H9NO2S	+	9.94
	5-Aminolevulinate	C5H9NO3	+	9.77
	gamma-L-Glutamyl-L-cysteine	C8H14N2O5S	+	10.05
	glutathione	C10H17N3O6S	+	10.05
	Acetylcholine	C7H15NO2	+	12.79
	L-Cystine	C6H12N2O4S2	+	10.84
	Homocystine	C8H16N2O4S2	+	0
	S-glutathionyl-L-cysteine	C13H22N4O8S2	+	0
	Mercaptoethanol	C2H6OS	+	21.5
	trypanothione	C27H49N9O10S2	+	10.46
	Glutathione disulfide	C20H32N6O12S2	+	11.05
	Bis-gamma-glutamylcystine	C16H26N4O10S2	+	11.32
	Benzoate	C7H6O2	-	6.41
	acetylcysteine	C5H9NO3S	-	7.11
	D-Glucono-1_4-lactone	C6H10O6	-	9.26
	Acetoacetate	C4H6O3	-	7.55
	Maleic acid	C4H4O4	-	9.86
	D-Arabinose	C5H10O5	-	10.42
	D-Fructose	C6H12O6	-	10.34

Mix	Compound Name	Formula	Polarity	Expected RT
3	Orthophosphate	H3O4P	-	10.64
	NADH	C21H29N7O14P2	-	9.44
	L-Dehydroascorbate	C6H6O6	-	11.81
	N-acetyl-L-glutamate	C7H11NO5	-	9.99
	ascorbate	C6H8O6	-	10.64
	Itaconate	C5H6O4	-	10.64
	D-Fructose 6-phosphate	C6H13O9P	-	10.74
	Adenosine 2'_5'-bisphosphate	C10H15N5O10P2	-	10.68
	2-phospho-D-glycerate	C3H7O7P	-	11.11
	pyrophosphate	H4O7P2	-	11.49
	NADPH	C21H30N7O17P3	-	0

Table D1: Authentic standard mixes run with every metabolomics experiment

Appendix E

Chapter 5.2.2.: Table of secreted proteins from *T. b. brucei* GVR35, with 109 proteins significantly identified using MASCOT (p-value <0.05).

Accesion no. (TritrypDB)	ID	Matches	Score
Tb11.01.1290	14-3-3-like protein, putative [a,b]	10(6)	104
Tb11.02.4700	14-3-3-like protein, putative [a,b]	9(2)	44
Tb11.01.3020	40S ribosomal protein L14, putative	2(1)	32
Tb10.6k15.3340	40S ribosomal protein S24E, putative	2(0)	17
Tb10.70.3360	40S ribosomal protein S3a, putative [b]	5(2)	57
Tb11.02.1085	40s ribosomal protein S4, putative [a]	3(1)	33
Tb09.211.0110	60S ribosomal protein L10, putative	3(1)	42
Tb09.211.4550	60S ribosomal protein L12, putative	6(2)	44
Tb927.7.5170	60S ribosomal protein L23a	1(0)	24
Tb09.244.2730	60S ribosomal protein L5, putative	6(1)	63
Tb10.70.7010	60S ribosomal protein L9, putative	3(1)	33
Tb09.211.3180	6-phosphogluconate dehydrogenase, decarboxylating, putative	2(1)	29
Tb11.02.1120	adenylosuccinate synthetase, putative [b]	9(2)	57
Tb09.211.4460	ADP-ribosylation factor, putative [a]	4(0)	26
Tb927.3.3450	ADP-ribosylation factor-like protein 3A, putative [a]	1(0)	17
Tb927.1.2340	alpha tubulin [a,b]	108(84)	2274
Tb09.160.4590	arginine kinase [a,b]	20(5)	116
Tb10.70.3710	aspartate aminotransferase [a,b]	10(1)	37
Tb927.1.2330	beta tubulin [a]	48(16)	524
Tb927.7.4060	calpain-like cysteine peptidase, putative [a]	20(2)	48
Tb10.70.0830	clathrin heavy chain [a]	36(5)	132
Tb927.2.1560	cyclophilin type peptidyl-prolyl cis-trans isomerase precursor, putative	3(0)	14
Tb11.03.0250	cyclophilin a [a]	12(6)	226
Tb11.01.8470	dihydrolipoyl dehydrogenase [a]	2(0)	27
Tb10.70.5650	elongation factor 1-alpha [a,b]	37(4)	116
Tb11.01.4660	elongation factor 1 gamma, putative [a]	8(1)	28
Tb10.70.2650	elongation factor 2 [a,b]	58(33)	942
Tb10.70.4740	enolase [a,b]	135(59)	1528
Tb09.160.3270	eukaryotic initiation factor 4a, putative [a,b]	4(3)	91
Tb10.70.1370	fructose-bisphosphate aldolase, glycosomal [a]	7(4)	153
Tb11.02.5450	glucose-regulated protein 78, putative [a]	4(2)	53
Tb10.6k15.3850	glyceraldehyde 3-phosphate dehydrogenase, cytosolic [a,b]	4(1)	58
Tb09.211.3540	glycerol kinase, glycosomal [a]	13(4)	168
Tb927.8.3530	glycerol-3-phosphate dehydrogenase [NAD+], glycosomal [a,b]	7(0)	44
Tb11.01.3170	guanine nucleotide-binding protein beta subunit- like protein [a]	3(0)	25
Tb11.01.0120	haloacid dehalogenase-like hydrolase, putative [a]	5(2)	56
Tb11.01.3110	heat shock protein 70 [a,b]	113(41)	914
Tb11.01.3080	heat shock protein 70 [a,b]	19(7)	237
Tb927.7.710	heat shock 70 kDa protein	6(4)	145
Tb10.26.1080	heat shock protein 83 [a,b]	38(15)	371

Accesion no. (TritrypDB)	ID	Matches	Score
Tb10.389.0880	heat shock protein, putative [a,b]	14(3)	114
Tb927.3.5340	Hsc70-interacting protein (Hip), putative	5(2)	42
Tb927.7.1320	HSP10 10 kDa heat shock protein, putative	5(2)	77
Tb10.70.0280	HSP60 chaperonin Hsp60, mitochondrial precursor [b]	7(5)	116
Tb10.v4.0053	hypothetical protein	29(13)	280
Tb10.70.1130	hypothetical protein	7(4)	150
Tb927.2.2550	hypothetical protein, conserved	2(0)	23
Tb927.4.2740	hypothetical protein, conserved	6(1)	78
Tb11.01.5680	hypothetical protein, conserved	6(3)	77
Tb09.160.5530	hypothetical protein, conserved	9(2)	72
Tb927.6.4140	hypothetical protein, conserved	1(1)	55
Tb11.01.1625	hypothetical protein, conserved	3(1)	51
Tb927.6.5000	hypothetical protein, conserved	13(8)	47
Tb10.389.0720	hypothetical protein, conserved	5(1)	37
Tb927.8.4820	hypothetical protein, conserved	3(1)	35
Tb11.47.0016	hypothetical protein, conserved	2(1)	35
Tb10.406.0610	hypothetical protein, conserved	9(2)	35
Tb09.211.3955	hypothetical protein, conserved	2(0)	34
Tb927.7.1420	hypothetical protein, conserved	120(1)	30
Tb09.211.1690	hypothetical protein, conserved	2(0)	27
Tb10.61.2620	hypothetical protein, conserved	2(0)	24
Tb927.6.3170	hypothetical protein, conserved	4(0)	24
Tb10.26.0680	hypothetical protein, conserved	6(0)	23
Tb927.1.4310	hypothetical protein, conserved	11(0)	22
Tb927.8.6330	hypothetical protein, conserved	15(0)	22
Tb11.01.4700	hypothetical protein, conserved	1(0)	18
Tb10.70.3070	hypothetical protein, conserved	1(0)	17
Tb927.3.3560	hypothetical protein, conserved	2(0)	16
Tb10.6k15.1500	hypothetical protein, conserved	1(0)	15
Tb10.61.0540	hypothetical protein, conserved	2(0)	13
Tb09.160.0465	hypothetical protein, conserved	3(0)	13
Tb10.6k15.0210	hypothetical protein, conserved	2(0)	14
Tb927.7.3440	I/6 autoantigen [a]	5(3)	68
Tb927.3.2960	inosine-adenosine-guanosine-nucleosidehydrolase [a]	5(2)	56
Tb11.02.2260	kinesin, putative	5(1)	45
Tb10.61.1750	C-terminal motor kinesin, putative [a]	15(5)	118
Tb10.70.5840	major vault protein, putative	10(0)	20
Tb10.70.5250	metacaspase MCA4; cysteine peptidase, putative [a]	1(0)	26
Tb09.211.0120	nascent polypeptide associated complex subunit, putative [b]	1(1)	50
Tb11.01.7800	nucleoside diphosphate kinase [a]	19(2)	110
Tb927.3.4290	PFR1 73 kDa paraflagellar rod protein [a,b]	7(1)	40
Tb10.6k15.2620	2,3-bisphosphoglycerate-independent phosphoglycerate mutase [b]	4(0)	29
Tb10.70.5520	piwi-like protein 1	42(10)	75
Tb11.01.1680	polyubiquitin, putative [a]	12(4)	91
Tb927.7.4770	cyclophilin-type peptidyl-prolyl cis-trans isomerise [a,b]	4(3)	79
Tb10.100.0170	proteasome alpha 2 subunit, putative	1(0)	19

Accesion no. (TritrypDB)	ID	Matches	Score
Tb10.70.0850	proteasome alpha 1 subunit, putative [a]	2(0)	20
Tb10.100.0120	proteasome alpha 5 subunit, putative	2(0)	20
Tb11.02.4040	protein transport protein Sec31, putative	2(0)	15
Tb10.61.2680	pyruvate kinase 1[a]	8(5)	166
Tb11.01.1350	S-adenosylhomocysteine hydrolase [a,b]	11(6)	159
Tb10.70.7050	TCP-1-delta t-complex protein 1, putative	2(1)	32
Tb09.211.2570	TCP-1-eta t-complex protein 1, putative	3(0)	14
Tb927.8.3150	TCP-1-gamma t-complex protein 1, putative	5(0)	29
Tb10.6k15.2330	TCP-1-theta t-complex protein 1, putative	3(0)	26
Tb11.02.0750	TCP-1-zeta t-complex protein 1, putative	4(0)	18
Tb11.03.0670	transcription factor, putative	7(1)	29
Tb11.02.3210	triosephosphate isomerise [a,b]	6(2)	73
Tb927.3.3760	tryparedoxin [a,b]	22(3)	79
Tb09.160.4250	tryparedoxin peroxidase [a,b]	23(0)	32
Tb927.7.1120	trypanothione/tryparedoxin dependent peroxidase 1, cytosolic	5(1)	55
Tb10.406.0520	trypanothione reductase [a,b]	7(1)	41
Tb09.211.3610	ubiquitin-activating enzyme E1, putative [a]	6(0)	30
Tb11.02.0815	ubiquitin-conjugating enzyme [a,b]	12(8)	229
Tb927.5.4950	variant surface glycoprotein (VSG, atypical), putative	14(1)	31
Tb11.21.0005	variant surface glycoprotein (VSG, pseudogene), putative	2(0)	20
Tb10.v4.0136	variant surface glycoprotein (VSG, pseudogene), putative	7(0)	14
Tb10.v4.0124	variant surface glycoprotein (VSG, pseudogene), putative	9(0)	13
Tb927.6.5630	variant surface glycoprotein (VSG, pseudogene), putative	7(0)	13

Table E1: Secreted proteins from *T. b. brucei* strain GVR35 (obtained with FASP, 5.2.2.). Identifications and accession no. were matched by mascot to TritrypDB.

Matches indicated the number of peptides matched against the protein sequence with the number in brackets showing significant hits, indicating homology or identity. Score > 29 indicates homology or identity to compared protein sequences with p > 0.05.

[a] protein identified in *T. gambiense* (Geiger *et al.*, 2010)

[b] protein identified in *T. congolense* (Grébaut et al., 2009)